GENETIC DIVERSITY OF AFRICAN ISOLATES OF *TOXOPLASMA GONDII*

MOHAMMED HAMED ALRUHAILI

Ph.D. Thesis **2015**

GENETIC DIVERSITY OF AFRICAN ISOLATES OF *TOXOPLASMA GONDII*

MOHAMMED HAMED ALRUHAILI

School of Environment and Life Sciences
College of Science and Technology
University of Salford, Salford, UK

Submitted in Partial Fulfillment of the Requirements of the Degree of Doctor of Philosophy, August 2015

Table of contents

List of Tables

List of Figures

Acknowledgments

With grateful thanks

First of all, I would like to thank Allah, whom I must thank for everything great in my life.

To the supervisor, Professor Judith Smith, for her generous help and her encouragement and guidance throughout the whole PhD period and through the process of writing this thesis.

To the co-supervisor, Professor Geoff Hide, for his co-operation and for all his help.

To Dr. Rachel Brenchley, for her great help, co-operation and useful advice.

To my mother, my wife and my both daughters for their outstanding assistance, motivation and love.

Declaration

No portion of this work referred to in this report has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Abbreviations

BLAST	basic local alignment search tool
BWA	Burrows Wheeler Aligner
bp	base pairs
DMEM	Dulbecco's modified Eagle's medium
EDTA	ethylene diamine tetra-acetic acid
FCS	foetal calf serum
GPI	glycophosphatidyl inositol
GRA	dense granule protein
GSCID	Genomic Center for Infectious Diseases
HFF	human foreskin fibroblast
IGV	Integrative Genomics Viewer
JCVI	J. Craig Venter Institute
kb	kilo base
Mbp	mega base pair
MDBK	Madin-Darby Bovine Kidney
MEGA	Molecular Evolutionary Genetic Analysis
MLEE	multilocus enzyme electrophoresis
MLST	multi-locus sequencing typing
MS	microsatellites
NGS	next generation sequencing
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PVM	parasitophorous vacuole membrane
RFLP	restriction fragment length polymorphism

RON	rhoptry neck protein
ROP	rhoptry protein
QC	quality control
SAG	surface antigen
SAM	Sequence Alignment Map
SNP	single nucleotide polymorphism
SRS	SAG1-related sequence
T. gondii	*Toxoplasma gondii*
TIGR	the Institute for Genomic Research
VCF	Variant Call Format
VSG	variable surface glycoprotein

Abstract

Toxoplasma gondii is an intracellular protozoan parasite and has the ability to infect all warm-blooded animals, including humans. While the three clonal lineages of *T. gondii* (I, II and III) predominate in North America and Europe, strains from other regions in the world appear to have more diverse genotypes. The aim of the current research is to analyse the level of genetic variation among local African *T. gondii* isolates in relation to their phenotype (genotype phenotype relationships). In this study, multi-locus nested PCR sequence analysis of seven Ugandan *T. gondii* isolates was applied using nine different genetic markers distributed across seven chromosomes and the apicoplast genome of *T. gondii,* which improved the discrimination power to detect variation among the local Ugandan strains. Although these markers were sufficient to separate global variation between *T.gondii* strains, they were not adequate to totally resolve within closely related local isolates. To understand the impact of local variation on strain diversity, whole genome sequence was generated for two Ugandan strains type II using Illumina MiSeq paired-end sequencing, revealing variations between these strains and the type II reference strain of *T. gondii* (TgME49). In this study, we have perhaps the first example of the deeper sequencing of isolates from the same geographical region at the same time point, which showed that they are non-identical. Novel polymorphisms were identified in a virulence associated gene in both Ugandan strains resulting in modification of the protein structure of this gene which could be associated with phenotype variation in the *in vitro* growth rate of these strains. Comparing the *in vitro* growth rates of the sympatric Ugandan strains, a cluster of 3 strains had higher growth. These were genotypically identical by using PCR sequencing technique, while the non-identical sympatric strains had lower growth rates, providing evidence that genotype may influence phenotype. An important finding was evidence of recombination between type II and III within three Ugandan strains, revealed through multi-locus PCR sequencing, and in an additional Ugandan strain through deeper whole genome sequencing. Six polymorphic markers were identified via analysis of three biologically relevant genes families (SRS, ROPs and GRA), enhancing the resolution power to identify variations among local type II strains of *T. gondii*. It is recommended that further study of these polymorphic markers is carried out and that they are added into the MLST analysis of *T. gondii*, especially between closely related local isolates.

1. CHAPTER 1: Introduction

1.1. The parasite, *Toxoplasma gondii*

Toxoplasma gondii is an intracellular protozoan parasite which is considered to be one of the most successful parasites due to its worldwide distribution and wide range of hosts. It has the ability to infect all warm-blooded animals including humans (Dubey, 2007; Dubey et al., 1998).

1.1.1. Life Cycle

The life cycle of *Toxoplasma gondii* has two sub-cycles, sexual and asexual cycles (Figure 1.1) (Dubey, 1998a; Dubey et al., 1998; Frenkel, 1973; Hutchison, 1965). The sexual cycle of the parasite takes place only in feline species; all species of domestic and wild cats can act as the definitive hosts. Infection of cats usually occurs orally as a result of ingestion of infected tissues. After several rounds of amplification of the parasite within gut cells, differentiation of both male and female gametes takes place and fusion of gametes occurs in the cat intestinal epithelium, leading to formation of fertilized oocysts (Hutchison et al., 1971). Oocysts are excreted into the environment in cat faeces and the process of meiosis begins, producing eight haploid progeny or sporozoites contained within a double-walled oocyst (Dubey and Frenkel, 1972). These oocysts persist in the environment and are highly infectious to intermediate hosts through the oral route (Dubey, 2007; Dubey, 2009a; Dubey et al., 1998).

The asexual cycle can occur in any warm-blooded animal, including birds, as intermediate hosts (Dubey, 2009b). After ingestion of *T. gondii* by the intermediate host either from drink or food contaminated with oocysts or tissues of infected animals, sporozoites are released from oocysts or bradyzoites from tissue cysts. These stages of *T. gondii* penetrate intestinal tissues, undergo the process of differentiation into the rapidly dividing tachyzoite inside the intestinal epithelium, and become distributed through the body via blood or lymph. After a few proliferation cycles, tachyzoites invade host cells and convert to slowly dividing bradyzoites. This leads to the formation of tissue cysts in a variety of tissues, such as neural, hepatic, pulmonary, cardiac or muscle cells during the chronic phase of the disease (Dubey, 2004). These tissue cysts can persist as dormant stages inside the host for life and are orally

infectious to definitive and other intermediate hosts (Dubey, 1997; Dubey et al., 1998; Hutchison et al., 1971; Sibley, 2011).

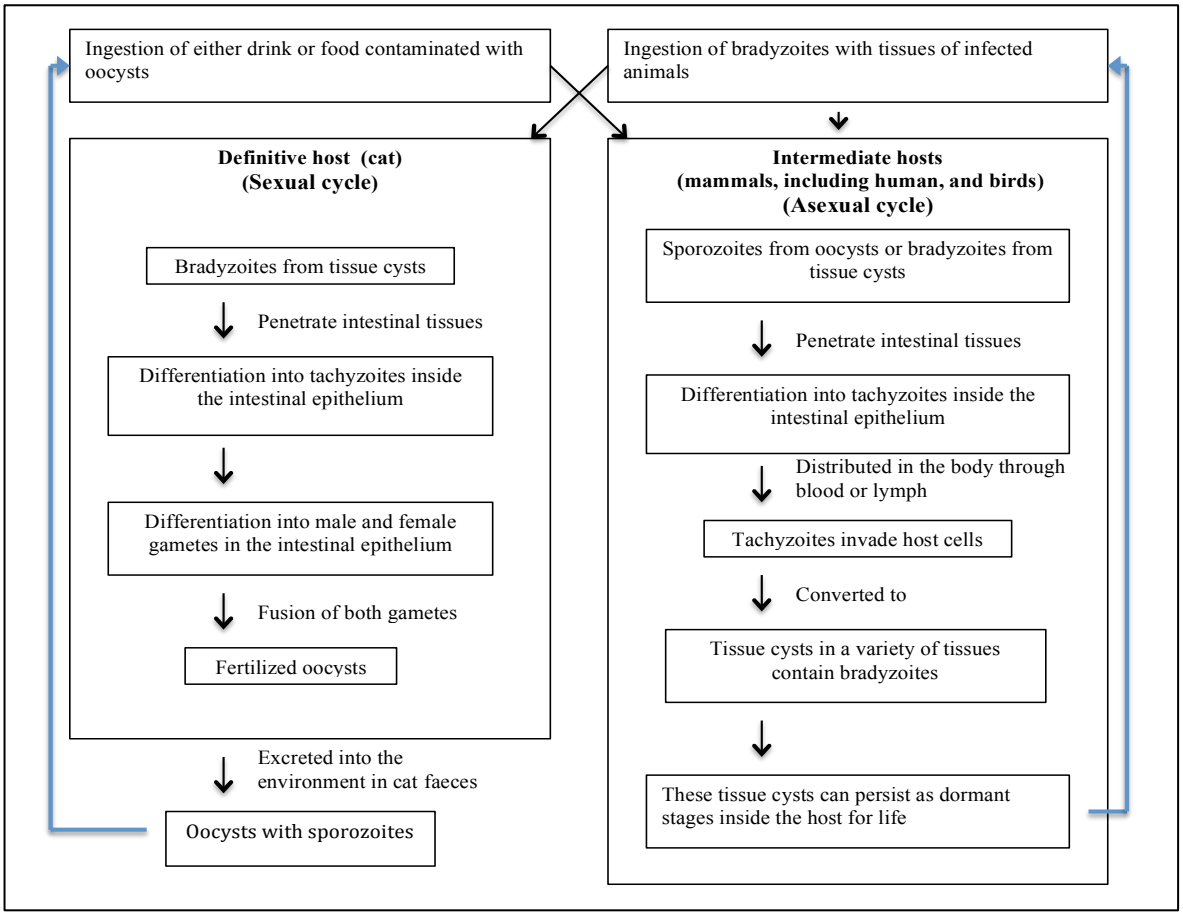

Figure 1.1:Life cycle of *Toxoplasma gondii*

1.1.2. Transmission

Humans can be infected with Toxoplasma through multiple routes, either from ingesting raw or undercooked meat containing tissue cysts or by ingesting water or food contaminated with oocysts (food-borne or water-borne) (Dubey and Jones, 2008). Interestingly, tissue cysts are more infectious for the definitive feline host than for intermediate hosts while, in contrast, oocysts are more infectious for intermediate hosts (Dubey, 2006). The extremely broad host range of the parasite can be explained in part by this variety in modes of transmission. *T. gondii* affects rodent behaviour by causing them to lose their fear response to cats, which remarkably enhances the transmission of infection (Webster, 2001). In addition, congenital infection is considered to be an important transmission route for *T. gondii* infection in most animal species and humans (Dubey, 2009a) which leads to transmission of infection from mother to fetus as the parasite crosses the placenta. In some hosts, there is evidence that

congenital transmission might occur through multiple generations as true vertical transmission, such as in sheep (Hide et al., 2009) and mice (Beverley, 1959).

The sexual and asexual cycles can produce three forms of the infectious stages of *T. gondii*, all of which are haploid. These stages are the production of tachyzoites, which are found during acute infection; bradyzoites, which are involved during chronic infection, and sporozoites, which occur within oocysts in the environment. Tachyzoites and bradyzoites have the ability to divide asexually, while sporozoites are produced as a result of meiosis of diploid oocysts after shedding and do not have the capacity to divide (Dubey, 1998b; Dubey et al., 1998).

The acute phase of *T. gondii* infection results from rapidly dividing tachyzoites which cause lysis and rupture of infected host cells and the emergence of parasites that can infect new host cells. Tachyzoites can be converted to slow dividing bradyzoites as a result of environmental stress, leading to formation of tissue cysts which remain dormant during the chronic phase. These tissue cysts turnover slowly, resulting in release of bradyzoites into the surrounding tissues. The subsequent inflammatory response inside the host can contain infection and prevent its reactivation. Reactivation of latent infection occurs in immunocompromised hosts, resulting in cyst rupture and conversion to tachyzoites. The flexibility of *T. gondii* for interconversion between these two different stages during its life cycle allows tissue cysts to infect intermediate hosts orally (Dubey, 1998a; Dubey et al., 1998).

In the definitive host, *T. gondii* can also differentiate into both stages of the asexual cycle, tachyzoites and bradyzoites, and this leads to the formation of infectious tissue cysts as in intermediate hosts (Dubey, 1997). The sexual cycle does not necessarily prevent the propagation of clones, as it has been shown that *T. gondii* lacks predetermined mating types, and the formation of both male and female gametes can occur from a single genotype which, by their fusion, results in progeny identical to the parent (Pfefferkorn et al., 1977). This is referred to as the selfing process. This process is thought to have significant implications for the epidemiology and the population biology of *T. gondii* (Wendte et al., 2010; Wendte et al., 2011).

1.2. The disease, Toxoplasmosis

1.2.1. Asymptomatic infection

Primary infection with *T. gondii* is asymptomatic in the majority of immunocompetent patients or is seen as a self-limiting disease (Hoyen, 1990). It is believed that infected humans remain infected for their whole lives, but parasite reactivation can occur as a result of immunosuppressive factors such as AIDS, or medication for inflammatory disease or transplantations (Porter and Sande, 1992).

1.2.2. Symptomatic infection in immunocompetent hosts

In immunocompetent patients, primary infections are mostly limited and manifest in fever, muscle weakness and cervical or occipital lymphadenopathy in the form of lymph node enlargement without tenderness or suppuration for about 4-6 weeks or even months in some cases (Ho-Yen, 2009).

1.2.3. Infection in AIDS and immunosuppressed patients

The opportunistic reactivation of a dormant *T. gondii* infection can occur when immunosuppression is initiated, as in AIDS patients or in transplant recipients as a result of immunosuppressive medication, leading to a fatal outcome as toxoplasmosis in this situation is considered to be a life-threatening disease. The association of HIV infection and CNS toxoplasmosis is well-established (Luft and Remington, 1992). It has been shown that the existence of encysted bradyzoites in the CNS of many AIDS patients is associated with high incidence of toxoplasmic encephalitis in these individuals (Denkers and Gazzinelli, 1998). In the early days of HIV emergence, prior to the development of anti-viral therapy, it was found that toxoplasmic encephalitis developed in about 10-50% of AIDS patients with chronic *T. gondii* infection (McCabe and Remington, 1988). CNS lesions are common in AIDS patients, and are caused by tissue destruction as a result of tachyzoite proliferation rather than immunopathologic effects caused by inflammatory response (Denkers and Gazzinelli, 1998).

In addition, the parasite can reactivate in the eye causing retinochoroiditis, or in the lung causing pneumonitis and acute respiratory failure (Ho-Yen, 2009; Luft and Remington, 1992).

1.2.4. Ocular disease

Ocular toxoplasmosis is one of the most common sequelae of chronic toxoplasmosis and presents in the form of retinochoroiditis. It can occur in congenitally or post-natally acquired infection resulting from acute infection or reactivation of dormant infection. However, ocular toxoplasmosis is more commonly associated with congenital infection compared to acquired infection (Brezin et al., 1994). The eye can be reached by *Toxoplamsa gondii* via the bloodstream in the form of free tachyzoites or as tachyzoites that exist in circulating leukocytes (Roberts and McLeod, 1999). These establish in the retina and form cysts. Pathology occurs due to the release of tachyzoites as a consequence of the rupture of tissue cysts which result in invasion and inflammation of the retina. Reactivation of infection and retinal disease in individuals with acquired, rather than non-congenital, toxoplasmosis is associated in some cases with reduction of immunity (Holland et al., 1988; Nicholson and Wolchok, 1976). In addition, it has been suggested that ocular toxoplasmosis can occur through transmission from the brain to the eye through the optic nerve (Mets et al., 1996). Furthermore, the development of tissue cysts that contain bradyzoites inside the eye can be involved in the reactivation of toxoplasmosis and is considered to be a significant feature in toxoplasmic retinochoroiditis pathogenesis which causes acute inflammation and can result in retinal scars that might cause blurred vision or blindness (Roberts and McLeod, 1999). Ocular infection can be manifested as pain, tearing, photophobia and finally loss of vision (Holland, 2003; Holland, 2004; Ho-Yen, 2009; Montoya and Remington, 1996).

1.2.5. Congenital toxoplasmosis

T. gondii infection can be transmitted congenitally to the foetus if the mother acquires the infection during pregnancy, as parasites cross the placenta and infect the foetus. Infection during the third trimester is considered to be the highest risk for congenital transmission of infection, while the first trimester is the lowest risk (Dunn et al., 1999). Without treatment of the mother during pregnancy, the incidence of acquired foetal infection during the first trimester is 10%-15%, the second trimester is 30% and the third is 60% (Wong and Remington, 1994). If the mother receives treatment with spiramycin, these incidences decrease to 4.5%, 17.3% and 28.9% respectively (Wong and Remington, 1994). The severity of congenital toxoplasmosis is also associated with the time of infection. Clinical manifestations are most severe when the infection is acquired during the first trimester of pregnancy compared with the second or third trimester (Dunn et al., 1999; Kravetz and

Federman, 2005). Additionally, it has been shown that there is an association between the severity of the disease and the density of parasites in amniotic fluid (Romand et al., 2004). Congenital infection can occasionally result from reactivation of infection in immunosuppressed women if acquired before pregnancy (Wong and Remington, 1994).

The variety of manifestations of congenital infection that occur in the foetus and in infants include spontaneous abortion, still-birth, a live infant with classic signs of congenital toxoplasmosis such as hydrocephalus or microcephalus, cerebral calcifications, mental retardation, seizures and retinochoroiditis (Hill and Dubey, 2002). The majority of cases are asymptomatic at birth, but most will develop neurological or ocular manifestations later in their lives. None of these signs is pathognomonic for toxoplasmosis and are similar to many infections causing congenital diseases (Ho-Yen, 2009).

1.3. Diversity and distribution of *Toxoplasma gondii*

1.3.1. Clonal strains of *Toxoplasma gondii*

T. gondii has a well-characterised sexual cycle which takes place in cats. Therefore, it was expected that the parasite would have a high genetic diversity among its isolates. However, analyses conducted on 35 isolates from North America and Europe by using isoenzymes (Darde et al., 1992) and PCR-restriction fragment length polymorphism (PCR-RFLP) (Howe and Sibley, 1995; Sibley and Boothroyd, 1992) all show the existence of extremely low genetic diversity (Darde et al., 1992; Howe and Sibley, 1995). These results suggested the presence of a clonal population structure where the majority of isolates (>94%) fell into three main clonal types, referred to as types I, II and III. A study by Howe and Sibley, (1995) estimated that the percentage of recombinant strains which represent mixtures of the main three clonal strains was low at approximately 5% (Grigg and Suzuki, 2003; Howe and Sibley, 1995).

The main clonal lineages of *T. gondii* were later shown to have arisen by recombination. By sequencing analysis of 18 different polymorphic genes of the main three clonal strains, it was shown that the major *T. gondii* strains are restricted to only two allelic types that are randomly located across the three strains (Grigg et al., 2001a). This restricted dimorphic gene pool suggested that these clonal lineages originated from one common ancestor as recently as 10000 years ago, and that this expanded rapidly to populate a range of hosts (Grigg and Suzuki, 2003; Su et al., 2003). These two allelic types were referred to as A

(Adam) and E (Eve) (Meisel et al., 1996). Analysis of exotic strains showed that they also carry A or E alleles within sequences at many loci (Boothroyd and Grigg, 2002; Grigg et al., 2001a). Khan et al. (2005) demonstrated that each locus in the three main clonal strains consisted of only 2 alleles by mapping the *T.gondii* genome to about 300-kb intervals across the 14 chromosomes with 250 SNP markers.

The existence of the sexual transmission of *T. gondii* in cats seems to have little impact on the clonality seen in this parasite. This could be explained by two factors. Firstly, the non-obligatory nature of the sexual cycle for transmission. Transmission can occur through asexual cycles between different intermediate hosts and its reproduction can be efficient without the necessity for a sexual cycle. Secondly, it is known that this parasite is haploid; while the infection of the definitive host occurs with only one strain, the production of the infectious oocyst will be genetically identical to the original infecting strain. Moreover, although a cat can be infected with multiple strains, through ingestion of an intermediate host infected with different strains types, this will be a rare event as the frequency of naturally acquired mixed infection is low among intermediate hosts (Boothroyd and Grigg, 2002).

1.3.2. Atypical strains of *Toxoplasma gondii*

While the three clonal lineages predominate in North America and Europe, strains from other regions in the world appear to have more diverse genotypes. By analysing isolates from South America, Asia and Africa by using PCR-RFLP or microsatellite markers, it was revealed that the majority of these isolates have type I, II or III alleles which are identical to those in the main three lineages, but some novel alleles are also detected. These isolates were classified as either recombinant or atypical strains. The recombinant genotypes have mixtures of classical alleles, while atypical, unusual, non-archetypical or exotic strains have a variety of novel alleles and unique polymorphisms (Darde, 2008).

Atypical (exotic) strains are characterized by the existence of many unique polymorphisms and novel alleles (Ajzenberg et al., 2004; Grigg et al., 2001a; Su et al., 2003). Phylogenetic analysis, based on microsatellites, suggests that these atypical genotypes are disseminated across the tree with no clear structure, or association with the main three lineages (Ajzenberg et al., 2004). Although there is clear divergence among these strains, the overall level of

sequence polymorphism is modest (Ajzenberg et al., 2004; Grigg et al., 2001a; Su et al., 2003).

The first atypical strain was identified in France. MAS was isolated from a human congenital toxoplasmosis case, while the second strain, CASTELLS, was isolated in Uruguay from an aborted sheep (Ajzenberg et al., 2004; Grigg et al., 2001a; Su et al., 2003). In addition, an atypical strain was isolated from marine mammals (referred as type X) (Conrad et al., 2005; Miller et al., 2004), a cougar isolate collected from Canada (Grigg et al., 2001a; Lehmann et al., 2000; Su et al., 2003), nine isolates from French Guiana (Ajzenberg et al., 2004; Carme et al., 2002), and another isolate from a human congenital toxoplasmosis from France (Ajzenberg et al., 2004).

1.3.3. Geographical variation of strains

Many published studies have revealed that type II strains are the predominant type found in North America and Europe (Table 1.1). In France, type II is associated with about 90% of human congenital toxoplasmosis and also all cases which were derived from animals (Ajzenberg et al., 2002b). Some studies revealed that type I and II are more frequent in Spain, while type II is dominant in Portugal (Fuentes et al., 2001; de Sousa et al., 2006). In the USA, there is also a predominance of type II in wildlife (Dubey et al., 2004a).

Isolates from South America are more genetically divergent compared with those from North America and Europe. It seems that the common clonal lineages which differ significantly from the main three types are circulating on this continent (Ajzenberg et al., 2004; Carme et al., 2002). In Asia, it has been shown that strains have more limited genetic diversity compared to South America (Dubey et al., 2005b). Isolates from Cameroon in Africa analysed via microsatellites revealed the existence of fixed combination of type I and III alleles which suggest a unique clonal African type (Ajzenberg et al., 2004).

Table 1.1: Geographical variation of *Toxoplasma gondii* strains.
MS: microsatellites, RFLP: restriction fragment length polymorphisms, SNP: single nucleotide polymorphisms

Origin	Host	Method	No. of isolates	Results				Reference
				Type I	Type II	Type IIII	Atypical	
USA	Pigs	RFLP, MS	25		20	5		(Lehmann et al., 2003)
USA	Wildlife	RFLP	43	3	35	5		(Dubey et al., 2004a)
Spain	Human	RFLP	25	10	10	5		(Fuentes et al., 2001)
Portugal	Pigs	RFLP, MS	15		10	4	1	(de Sousa et al., 2006)
France	Human	MS	86	7	73	2	4	(Ajzenberg et al., 2002b)
French Guyana	Human	SNP & MS	13	1		1	11	(Ajzenberg et al., 2004)
French Guyana	Humans	MS	16				16	(Carme et al., 2002)
India	Chickens	RFLP, MS	7		2	5		(Sreekumar et al., 2003)
Sri Lanka	Chickens	RFLP	12		6	6		(Dubey et al., 2005b)
Africa	Human	SNP & MS	3		1		2 (recombinants)	(Ajzenberg et al., 2004)
Egypt	Chickens and ducks	RFLP	9		3	17		(Dubey et al., 2003a)

1.3.4. Methods for monitoring strain diversity

The first evidence of the existence of discrete strains of *Toxoplasma gondii* in nature was the analysis of isoenzyme profiles for 35 isolates of this parasite from Europe and North America which suggested the presence of diversity among these strains, with most strains clustering into a small number of zymodemes (Darde et al., 1988). This was subsequently confirmed by genetic analysis, restriction fragment length polymorphisms (RFLP), and it was concluded that the majority of strains fall into one of three genotypes, types I, II and III (Howe and Sibley, 1995; Sibley and Boothroyd, 1992). Genetic studies of the clonal types revealed the limitation of the allelic polymorphism in the major *T. gondii* strains to only two allelic types, called A and E, to highlight the fact that these clonal strains have arisen from two distinct originators by sexual recombination (Boothroyd and Grigg, 2002; Grigg et al., 2001a).

From the genetic studies on Toxoplasma isolates it is possible to detect recombination. This can be done by mapping the pattern of alleles on chromosomes and analysing them to reveal whether exchange has occurred. The presence of natural recombinant genotypes in *Toxoplasma gondii* (Bontell et al., 2009) supported the occurrence of a sexual cycle in this parasite (Grigg and Suzuki, 2003). However, the predominance of the main three lineages over other less successful recombinants in nature could be explained by recent clonal expansion of biologically successful lineages originating from one common ancestor as recently as 10000 years ago, as described earlier (Grigg and Suzuki, 2003; Su et al., 2003). It has been suggested that this is the consequence of a selective advantage which presents in these three types, perhaps through affecting their biological characteristics such as colonization or transmission within a particular host (Grigg and Suzuki, 2003).

As the geographical distribution of studies broadened, it was shown that isolates collected from tropical areas such as Africa, Reunion Island, the Caribbean (Ajzenberg et al., 2004), and Brazil (Ferreira et al., 2006), were characterised by mixed genotypes of the three main clonal strains. Also, a few isolates with mixed genotypes have been identified in samples from wildlife (deer and bears) in North America (Ajzenberg et al., 2004; Howe and Sibley, 1995). In North America and Europe, recombinant strains were very rare in isolates collected from the domestic population of *T. gondii*. For instance, two recombinant strains were isolated from AIDS patients in the USA (Howe and Sibley, 1995); four from human congenital toxoplasmosis in France (Ajzenberg et al., 2002b; Ajzenberg et al., 2004); and five from human ocular toxoplasmosis in North America (Grigg et al., 2001b). Regions bearing the typical dimorphic allele patterns have been found in several loci of atypical strains (Su et al., 2003). This finding supports the occurrence of crosses between these strains and the clonal lineages (Sibley, 2003).

The effects of genetic exchange can be shown as reduction in linkage disequilibrium, modification of genotypes and decrease in phylogenetic divergence (Tibayrenc and Ayala, 2002). It is suggested that the role of sexual recombination may be reduced in *Toxoplasma*, as occurs with many parasites such as *Leishmania, Trypanosoma cruzi or Entamoeba histolytica* (Tibayrenc et al., 1990). Lehmann et al. (2004) compared data from microsatellites and found that isolates derived from Brazilian chickens had lower linkage disequilibrium than strains isolated from North American domestic animals. This assumes a higher frequency of outcrossing in Brazil. Ferreira et al. (2006) identified recombinant

genotypes in all 20 Brazilian isolates by PCR-RFLP at 8 different independent loci, again confirming the significant role of sexual recombination in these strains.

Additionally, it was revealed by Ajzenberg et al. (2004) that analysis of 21 domestic and 22 unusual strains (i.e. strains with different host or geographical origin) by using sequences of microsatellites markers detected recombinant and atypical strains which had a higher genetic diversity compared to the typical clonal strains. The weak relationship among these isolates is shown by phylogenetic reconstruction; it differs from the typical structure of a clonal organism like *Leishmania, Trypanosoma cruzi or Entamoeba histolytica* (Banuls et al., 1999; Tibayrenc et al., 1990). Phylogenetic and genetic analyses show that *T. gondii* has a complex population structure which is characterized by a mixture of sexual and asexual propagation. This trait is shared by other parasites in the Apicomplexa phylum such as *Cryptosporidium* (Mallon et al., 2003).

It is possible that the level of recombination among Toxoplasma strains may depend on diversity of habitat and availability of felid hosts. Thus, Toxoplasma may have higher rates of sexual recombination between isolates in wildlife (e.g. French Guiana) (Table 1.1). The predominance of archetypal strains (I, II and III) initially seen in Europe and North America may be associated with the low number of wildlife samples. In North American and European strains there is strong evidence for domination of clonal propagation and domestic transmission. However, spread of this pattern to include wildlife species in these areas is still not well-defined due to sample bias. Furthermore, it is difficult to understand the patterns of transmission of this parasite and its genetic exchange beyond these regions. By observing the distribution of recombinant and atypical strains, it was suggested that these strains are located in environments which are less dominated by domesticated species (e.g. Africa, Brazil, the Caribbean islands). The higher levels of allelic variation in atypical strains suggest they have a more diverse population structure, with a wider range of genotypes maintained by recombination of these strains in this species-rich environment. Although atypical strains are genetically different from the three main lineages, such as the cougar (Lehmann et al., 2000) and French Guianan strains (Ajzenberg et al., 2004), they could represent unknown clones that are adapted to wild animals. To understand this issue more clearly, intensive epidemiological sampling of wildlife habitats would be required. The new clonal types, if they exist, should be distinguished from transient clones by widespread distribution over a broad geographical and host range, and persistence over a long period of

time (Tibayrenc et al., 1990).

1.3.5. Geographical variation and local adaptation

It is likely that only a few host species are involved in the domestic cycle of *T. gondii*, including cats, birds and a few meat-producing animals. This may have contributed to the reduction in complexity of the genetic pool of the parasite in the domestic cycle. It would seem that the main three clonal lineages have been the most successfully adapted genotypes to these domestic hosts (Lehmann et al., 2003). In terms of timing it is thought that strains diverged some 10000 years ago, corresponding with the domestication of companion and agricultural animals (Su et al., 2003). Through the domestication of cats and the expansion of a limited range of domestic meat and milk-producing animals in North America and Europe, these parasite lineages would have come to dominate the agricultural habitat. Currently, domination by these major clonal lineages is maintained through the trade of food animal products and in human travel between different countries. In addition, transmission of clonal genotypes can spread to the neighbouring wild environment from farms which are considered as reservoirs of infection (cats, rodents and birds) (Lehmann et al., 2003), resulting in the reduction of genetic diversity. In Europe and North America, landscape change as a result of human activity may favour genetic drift for evolution of Toxoplasma by reducing recombination and consequently genetic flow in *T. gondii*. Therefore, this would diminish the adaptive potential of *T. gondii* in a domestic environment, and thus its biodiversity.

If it is true that *T. gondii* strains have co-evolved with certain host species, then it may also be true that the diversity and density of host species are associated with the diversity of the parasite. High diversification of parasite genotypes would therefore be found in species-rich habitats, such as the rainforest, which contain different species of mammals and birds. There is evidence that certain *T. gondii* strains in the wild environment in some parts of the world (e.g. South America) are different from the main clonal lineages and have more diverse genotypes. In reality, it is difficult to judge whether this is due to biogeographical separation or whether they represent ancient populations. Additional sampling and genotyping of *T. gondii* strains in these areas is needed to estimate gene flow in these populations, and reveal whether they do make more use of the sexual cycle or resolve into a further complex of the clonal lineages with infrequent recombinants. The frequency with which recombinant strains arise is based on the degree of genetic diversity and the transmission rate to felids. If both

these factors are high, this would lead to the increasing likelihood of recombination and mixed infections. For example, in France, while type II is the predominant type, if a cat ingests two different strains during a short period of time, recombination would mainly occur between two type II isolates, producing a type II oocyst. In environments such as Brazil, recombination between different genotypes is more likely to produce a mixed genotype oocyst. Additionally, it has been shown that experimental re-infection of an intermediate host occurs when the genotype of the second infecting isolate is different from the first one (Dao et al., 2001). Therefore, re-infection is higher in wild intermediate hosts due to the greater genetic diversity, resulting in a greater occurrence of mixed infections in the wild environment where both factors that affect genetic variation - recombination and genetic flow - are high, which allow *T. gondii* to preserve an optimal capacity for adaptation.

1.4. Virulence/phenotype

1.4.1. Virulence in humans

In North America and Europe, most cases of human toxoplasmosis in AIDS and congenital infections are associated with type II strains (Ajzenberg et al., 2002b; Howe and Sibley, 1995; Howe et al., 1997). However, a study in Spain reported dominance of type II strains in AIDS patients, while type I strains were associated with congenital infections (Fuentes et al., 2001). Another study from the USA revealed the relationship between severe ocular toxoplasmosis in immnunocompetent patients and type I strains and new recombinant genotypes (Grigg et al., 2001b).

1.4.2. Virulence of the main three clonal lineages

The severity of *T. gondii* acute infection is considered to be one of the most remarkable phenotypes among *T. gondii* different strains. It was recognised, from a broad survey of natural isolates, that 18 strains of *T. gondii* with a type I genotype all showed high virulence during acute infection in mice (Howe and Sibley, 1995; Sibley and Boothroyd, 1992). The acute virulence phenotype is well-defined in the mouse model. Type I strains have a lethal dose (LD100) (the minimal dose which causes 100% mortality) of a single infectious organism, while mouse-avirulent strains type II and III have a lethal dose of 10^3 infectious organisms and normally generate chronic infection in the mouse (Sibley et al., 2002). There is some evidence that type I isolates are associated with severe pathology in humans. For instance, a study in the USA revealed the relationship between severe ocular toxoplasmosis

in immnunocompetent patients and type I strains (Grigg et al., 2001b). However, epidemiological studies have shown that type I strains are rare in human and animal infections, and type II strains are considered to be the most common source of human toxoplasmosis (Howe and Sibley, 1995). During human toxoplasmosis, it is considered that active proliferation of the tachyzoite stage during the life cycle is the main cause of pathology. Thus, understanding the basis of this acute virulent phenotype is essential to identify the basis of pathology and consequently to improve interventions against toxoplasmosis (Sibley et al., 1999).

It is possible that the acute virulence of the RH strain, commonly used in experiments, could have resulted from widespread laboratory passages since its first isolation (Sabin, 1941). However, more than 20 isolated type I strains which share a genotype also exhibit the acute virulence phenotype in mice, indicating the significant correlation between acute virulence phenotype and genotype of the parasite (Sibley and Boothroyd, 1992).

1.4.3. Virulence of atypical and recombinant lineages

In atypical and natural recombinant strains, it is difficult to predict the correlation between phenotype and genotype of the parasite (Darde, 2008). Depending on their different inherited combination of genes, the phenotypes of these strains range from highly virulent to intermediate or non-virulent types (Grigg and Suzuki, 2003). These strains have been classified into three groups, based on their virulence in terms of mortality in mice within four weeks of infection. The classification is virulent (100% death of mice within four weeks); intermediate virulent (30% - <100% mortality); or non-virulent (<30% mortality) (Pena et al., 2008). An experimental cross between two avirulent strains types II and III resulted in progeny that were either avirulent, similar to their parents, or highly virulent (Grigg et al., 2001a). In addition, significant differences in the dissemination patterns in mice were found in virulent progeny that resulted from cross between type II and type III avirulent strains (Saeij et al., 2005). Additionally, another example of cross between a highly virulent strain (type I) and low virulent strain (type III) revealed that host mortality caused by new progeny ranged from low to 100% (Taylor et al., 2006).

1.4.4. Migration and virulence

T. gondii is considered to be a successful organism because it has the ability to cross biological barriers such as blood-brain barrier, placenta, or gut epithelium (Saeij et al., 2005).

Barragan and Sibley (2002) revealed that type I strain isolates exhibit high migration measured by *in vitro* assays, compared with other clonal types II and III. Barragan and Sibley (2003) developed an *in vitro* assay of migration by using polarized host cell monolayers, while active parasite motility was required for transmigration. It was shown that transmigration across polarized Madin-Darby canine kidney (MDCK) cells were increased in type I strains, compared with type II and III strains. In addition, the long–distance migration (LDM) phenotype was demonstrated by several type I strain parasites. *In vivo*, it was shown that RH type I strain parasites migrate to the spleen (LDM) more successfully than type II and III parasites. An analysis of type I/III cross recombinant progeny showed that there is an association between migratory capability (LDM phenotype) and a locus on chromosome VIIa, which mapped to the same chromosome region previously linked to acute virulence (Barragan and Sibley, 2003).

1.4.5. Growth rate and virulence

In protozoan pathogens, growth rate is considered to be an important virulence trait (Taylor et al., 2002). In *T. gondii*, the association between growth rate and virulence has been widely described (Kaufman et al., 1959). The burden of the *T. gondii* parasite plays a significant role in pathogenesis in mice (Mordue et al., 2001) by overstimulation of the immune system (Nguyen et al., 2003) resulting in high level secretion of T helper cell type I (Th1) cytokines, leading to apoptosis and organ damage. In addition, while a single type I strain tachyzoite has the ability to produce high parasite loads and thus a high level of Th1 cytokines, the same level of cytokines and pathology can be generated by increased inoculation of type II strain parasites. This increased parasite burden could be explained by increased resistance to the host immune system, an inherited shorter doubling time (Td), and/or a lower conversion to the slow growing bradyzoite stage. Several studies revealed that type I strains grow more rapidly that other type II and III strains (Radke et al., 2001). However, it is essential to differentiate between Td, time to lysis of host cells, and the parasite burden in culture. Complete lysis of cultured cells by type I RH strains occurred more rapidly than by type II and III strains without a significant difference in Td between these strains (Sibley et al.,

2002). This inconsistency in growth rate could be explained by an elevated reinvasion rate of type I strain parasites (Saeij et al., 2005). In *T. gondii*, Td is related to the different developmental stages of the parasite. It was demonstrated by Jerome et al. (1998) that the growth rate of tachyzoites emerging from a cell infected with type III sporozoites was rapid for 20 divisions; thereafter their growth was slow. This decrease in the growth rate of these tachyzoites was associated with the appearance of bradyzoite markers such as bradyzoite antigen (BAG1). A spontaneous mutant of the VEG strain MS-J was observed to sustain a rapid initial growth rate *in vitro* without the appearance of any bradyzoite markers and its virulence was 1000 times that of the parental VEG strain.

1.4.6. Identification of virulence genes by linkage mapping

In *T. gondii*, the existence of a classical genetic linkage map provides an excellent opportunity to achieve classical genetic studies which aimed to locate unknown genes that control specific phenotypes (Sibley and Boothroyd, 1992). Experimental crosses between different strains with different virulence can provide a valuable approach to screening for virulence-associated loci. An experimental cross between two avirulent strains, type II strain ME49 and type III strain CEP, resulted in progeny that were either avirulent, similar to parents, or to highly virulent (Grigg et al., 2001a). In addition, significant differences in the dissemination patterns in mice were found in virulent progeny that resulted from cross between type II and type III avirulent strains (Saeij et al., 2005). In research to discover the genetic basis of acute virulence in *T. gondii*, a cross between a high virulent type I strain named GT1, which has a virulent phenotype similar to RH strain, and low virulent type III strain CEP, revealed that host mortality caused by new progeny ranged from low to 100% (Taylor et al., 2006). This study analysed the association between a locus that is linked to SAG3, which is located on chromosome XII, and acute virulence (Table 1.2). Furthermore, there are other candidate genes which are located on chromosome XII (ROP5 and Adenosine kinase) (Saeij et al., 2006). Another study of natural recombinants in *T. gondii* revealed the association between acute virulence phenotype with a locus linked to SAG1 on chromosome VIII (Howe et al., 1996). However, analysis of recombinant progeny resulting from type I/III cross showed no significant linkage with markers located on chromosome VIII. Instead, a significant linkage was recognized with chromosome VII (Sibley et al., 2002). Many virulence-associated genes are located on chromosome VIIa and encode molecules such as ROP18, a rhoptry protein of the ROP2 family, a highly polymorphic serine-threonine kinase which is secreted during *T. gondii* invasion. It is unlikely that virulence will be controlled by

a single gene. Another gene located on chromosome VIIb encodes ROP16, a rhoptry protein kinase which has a role in controlling the induction of interleukin 12 which is secreted by mouse macrophages (Saeij et al., 2006) (see Chapter 4). The application of these experimental crosses between different strains in naturally recombinant strains would be useful in the detection of the expression level of the virulence genes and hence the prediction of the virulence level in a certain isolate. An initial study carried out by Pena et al., (2008) to determine the virulence differences between atypical isolates revealed one of the candidate virulence gene found on chromosome VIIa was in close proximity to the ROP18 locus.

These preliminary studies have provided significant findings. Firstly, that virulence is a hereditary trait, as clones reveal constant virulent phenotypes related to parental line. Secondly, some clones exhibit intermediate virulent phenotype which is different from the parental phenotypes (virulent and non-virulent), indicating that acute virulence is multigenic (Sibley et al., 2002). However, it is not obvious from these studies if this linkage is due to the existence of a gene or genes in certain regions which contribute to acute virulence or whether this occurs accidentally (Sibley et al., 2002).

Table 1.2: Examples of virulence genes in *Toxoplasma gondii*

Virulence genes	Chromosomes
ROP18	Chromosome VIIa
ROP16	Chromosome VIIb
SAG3, ROP5 and Adenosine kinase	Chromosome XII
SAG1	Chromosome VIII

1.5. Aims of this thesis

There are many different strains of the parasite *Toxoplasma gondii* in the world, across Europe, Africa and USA. Three clonal strains predominate. Although they are co-distributed, it seems that recombination between isolates is very rare. In addition, the pathology caused by infection varies greatly with the strain type. The main focus of this research project was to investigate the level of genetic variation among *Toxoplasma gondii* isolates in relation to their phenotype (genotype-phenotype relationships).

The first aim was to seek evidence of local allelic variation and to test whether differences in the phenotype of strains relate to genetic variation at key loci within a collection of seven

isolates from Africa (Uganda) - TgCkUg1, 3, 5, 6, 7, 8 and 9 strains - which were subjected to intensive sequencing with more genetic markers. The next aim was to generate the whole genome sequence of two type II *T.gondii* strains (TgCkUg8 & 9) using the Illumina MiSeq Paired-End Sequencing and conducting a comparison with the type II reference strain (TgME49) through alignment. This enabled an investigation to estimate the level of local variation within type II strains and to search for further evidence of recombination in African strains other than TgCkUg2 (Bontell et al., 2009). This was followed by the search for novel allelic variants within three biologically relevant gene families for assessment of local strain diversity and to examine variation in these genes in relation to their biological functions.

2. CHAPTER 2: Multi-locus nested PCR sequence analysis of Ugandan *Toxoplasma gondii* strains

2.1 INTRODUCTION

2.1.1 Genetic diversity

The first evidence for the existence of discrete strains of *Toxoplasma gondii* in nature was the analysis of isoenzyme profiles for 35 isolates of this parasite from Europe and North America which suggested the presence of diversity (Darde et al., 1992). As discussed in Chapter 1 (Section 1.3.1), this was later confirmed by genetic analysis which implied that the majority of strains fall into one of three genotypes: types I, II and III (Howe and Sibley, 1995; Sibley and Boothroyd, 1992).

While the three clonal lineages predominate in North America and Europe, strains from other regions in the world appear to have more diverse genotypes (Sibley et al., 2009). Analysis of isolates from South America, Asia and Africa by using either PCR-RFLP or microsatellite markers, reveals that the majority of these isolates have type I, II or III alleles which are identical to those in the main three lineages, but some novel alleles were also reported. These isolates were classified as either recombinant or atypical strains. The recombinant genotypes have mixtures of classical alleles, while atypical, unusual, non archetypal or exotic strains have a variety of novel alleles and unique polymorphisms (Darde, 2008).

2.1.2 Genotyping methods for *Toxoplasma gondii*

Our understanding of variation is clearly dependent on the number and discrimination of markers available for analysis. Darde (2008), in reviewing the use of multilocus analysis with microsatellite and PCR-RFLP markers of isolates collected from South America, Africa and Asia, comments on the classification of new isolates which are different from the main three clonal types as either atypical or recombinant isolates. She concluded that this is relatively artificial, and that increasing the number and the discriminating power of genetic markers plays a significant role in genotyping. The use of additional genetic markers would have the potential to detect mixtures of classical type I-III alleles in recombinant strains or

may even detect unique polymorphisms in isolates which had been previously classified as one of the main clonal types. Several methods have been used for genotyping of *Toxoplasma gondii* isolates.

2.1.2.1 Multilocus enzyme electrophoresis (MLEE)

Enzyme electrophoresis techniques were used in early studies for typing of *Toxoplasma gondii*, using isoenzymes or isozymes. Isoenzymes are variable molecular forms of a specific enzyme which have variable migration speed in electrophoresis depending on their electric charges (Tibayrenc, 2009). These different electric charges of isoenzymes, which are proteins, originate from the charge of each amino acid in the protein. Therefore, the differentiation in migration speeds is directly related to the primary sequence of amino acids of each enzyme, and consequently to the genetic sequences of the genes that encode these proteins. The staining reaction based on the specific substrate of each enzyme makes it possible to show the activity of each enzyme, which reflects only one genetic locus (Tibayrenc, 2009). In an initial study, 4 isoenzymes in 7 strains collected from France were classified by this technique and were found to fall into three main zymodemes - Z1, Z2 and Z3 (Darde et al., 1988). The method was later applied to 35 isolates collected from Europe and North America using six different isoenzymes and it was found that they grouped into main five zymodemes - Z1, Z2, Z3, Z4 and Z5 (Darde et al., 1992). Later, 86 isolates from Europe, North and South America were classified into twelve zymodemes through application of different six isoenzymes (Darde, 1996) (Table 2.1).

Table 2.1: Examples of multilocus enzyme electrophoresis (MLEE) techniques for genotyping of *Toxoplasma gondii*

No. of markers	No. of isolates	Results (correspond to)				Reference
		Type I	Type II	Type IIII	Atypical	
4	7 From France	Z1 1	Z2 4	Z3 2		(Darde et al., 1988)
6	35 From Europe and North America	Z1 6	Z2 & Z4 24	Z3 4	Z5 1	(Darde et al., 1992)
6	86 From Europe, North and South America	Z1 17	Z2 & Z4 52	Z3 9	Z5 to Z12 8	(Darde, 1996)

These studies revealed the extent of variation in *Toxoplasma gondii* strains and showed that most isolates fell into a small number of zymodemes (Table 2.1). Some of the reference strains of *T. gondii* involved in these studies showed agreement of their zymodemes

classification with the classical typing using microsatellites (MS) and PCR-RFLP genotyping techniques (Table 2.2). The method is useful but has quite serious disadvantages for wider epidemiological studies due to the requirement for huge amounts of purified parasites to detect protein markers (Sibley et al., 2009).

Table 2.2: Examples of the reference strains genotyping of *Toxoplasma gondii* by multilocus enzyme electrophoresis (MLEE), Microsatellites (MS) and PCR-RFLP techniques

Reference strains	**Multilocus enzyme electrophoresis (MLEE)** (Darde, 1996)	**PCR-RFLP** (Howe and Sibley, 1995)	**Microsatellites (MS)** (Ajzenberg et al., 2004, Ajzenberg et al., 2002a)
RH	Z1	I	I
ME49	Z2	II	II
NED	Z3	III	III
CEP	Z3	III	III

2.1.2.2 Microsatellite (MS) markers

Microsatellites (MS), also known as short tandem repeats (STR) or simple sequence repeats (SSR), are short repeated segments of DNA which are composed of 1 to 6 base pair long units that are repeated in tandem (Field and Wills, 1996). MS are widespread in both prokaryotic and eukaryotic genomes and can be located anywhere in the genome, both in non-coding and protein-coding regions, and can be detected by PCR. However, the majority of MS are located in non-coding regions, either in the introns or in the intergenic sequence. It is this type of MS that is usually selected as genetic markers, as they are mostly assumed to evolve neutrally. MS are largely used as genetic markers to differentiate population structure, as they have found to be rapidly evolving compared to the rest of the genome. Thus, it is suggested that MS have a significant role in the genome evolution by generation and preservation of genetic variation (Kashi et al., 1997). In *Toxoplasma gondii*, Ajzenberg et al. (2002b) was one of the first to use this approach and developed 8 MS markers to genotype 86 *Toxoplasma gondii* isolates from European countries, mainly from France (Table 2.3). It was clear that the majority of European isolates were genotyped as type II strains. In addition, a further study applied to 61 *T. gondii* isolates collected from North America and Europe through the application of a single microsatellite marker showed that the majority of these isolates were grouped into type II strains (Costa et al., 1997). Ajzenberg et al. (2004) also genotyped 43 *T. gondii* isolates that had been sampled from North America, Europe, South America and Africa using 5 MS. This study revealed that the

vast majority of these samples were classified into atypical strains in addition to the detection of 4 recombinant strains among them. Another study that genotyped 16 strains from French Guyana through application of 8 MS markers showed that all these isolates were typed as atypical strains (Carme et al., 2002) (Table 2.3). Reference strains of *T. gondii* were incorporated in many of these studies and always showed agreement with their zymodemes, based on classification through MLEE or classical PCR-RFLP genotyping techniques (Table 2.2). Although MS markers are sensitive, reliable and highly polymorphic, they can be highly affected by contamination of host DNA, so while they can be easily employed on isolates they are difficult to use on tissue samples (Sibley et al., 2009).

Table 2.3: Microsatellite analysis method for genotyping of *Toxoplasma gondii*

No. of markers	No. of isolates	Results					Reference
		Type I	Type II	Type IIII	Atypical	Recombinant	
8	86 From European countries	7	73	2	4		(Ajzenberg et al., 2002b)
5	43 From North America, Europe, South America and Africa	3	6	7	23	4	(Ajzenberg et al., 2004)
1	61 From North America and Europe	6	54	1			(Costa et al., 1997)
8	16 From French Guyana				16		(Carme et al., 2002)

2.1.2.3 PCR- Restriction fragment length polymorphism (RFLP) analysis

The restriction length polymorphism (RFLP) method has been extensively used for *Toxoplasma gondii* genotyping. SAG2 gene polymorphism analyzed by PCR-RFLP was the first method used to distinguish the main three genotypes of *T. gondii* by using two restriction sites (Howe et al., 1997). This typing method, based purely on the SAG2 locus, was applied in many studies; however, the low resolution of the technique was later found to result in misidentification of recombinant or atypical strains. For example, genotyping of 14 isolates by SAG2 PCR-RFLP methods revealed their genotype as type I; however, 4 were found to exhibit a type II genotype by use of 8 different microsatellite markers (Ajzenberg et al., 2002a). Thus, it became clear that genotyping based on this single marker would be insufficient for discrimination of the three main clonal groups and have no power to detect recombinant or atypical strains. For these reasons, a combination of different loci (multilocus genotyping) was applied. For example, Su et al. (2006) conducted a study to

genotype 46 *T. gondii* isolates collected from Europe, North and South America by using a panel of 9 PCR-RFLP markers. Together, these markers have the ability to distinguish the main three clonal lineages from each other and from atypical strains. Another example of improved power to detect recombinant or atypical strains by multilocus genotyping is a study conducted by Pena et al (2008) to genotype 46 samples from Brazil using 10 different markers. This study showed that all 46 isolates were classified as atypical strains. In addition, all 16 isolates that had been sampled from Brazil were identified as recombinant strains by the PCR-RFLP genotyping method by using 8 different genetic markers (Ferreira et al., 2006). It is obvious from examples of the use of the PCR-RFLP method for genotyping of *Toxoplasma gondii* summarized in Table 2.4 that multilocus genotyping improved understanding of the distribution of clonal lineages, the existence of atypical or recombinant strains and major geographical differences.

Table 2.4: Examples of PCR-RFLP method for genotyping of *Toxoplasma gondii*

No. of markers	No. of isolates	Results					Ref.
		Type I	Type II	Type IIII	Atypical	Recombinant	
1 (SAG2)	68 From France	7	55	6			(Howe et al., 1997)
1 (SAG2)	25 From Spain	10	10	5			(Fuentes et al., 2001)
6	106 From, North America and Europe	15	50	24			(Howe and Sibley, 1995)
5	12 From North America	3	3	1		5	(Grigg et al., 2001b)
7	16 From North America and Europe	4	4	5	1	2	(Lehmann et al., 2000)
3	28 From North America, Europe, Australia and Brazil	10	6		9	3	(Sibley and Boothroyd, 1992)
9	46 From Europe, North and South America	4	3	3	36		(Su et al., 2006)
8	16 From Brazil					16	(Ferreira et al., 2006)
10	46 From Brazil				46		(Pena et al., 2008)
11	164 From Central and South America		4	20	140		(Rajendran et al., 2012)
8	14 From China				2	12	(Chen et al., 2011)

The clonal structure of the *T. gondii* population originally led to a typing strategy based on only one or two loci, mainly SAG2 and/or SAG1, which had the capability to distinguish the main three genotypes in many studies (Fuentes et al., 2001; Howe et al., 1997). This strategy led to mis-identification of atypical and recombinant strains. For example, genotyping of 14 isolates by SAG2 PCR-RFLP methods revealed their genotype as type I; however, 4 were found to exhibit a type II genotype by using 8 different microsatellite loci of 6 isoenzymatic markers (Ajzenberg et al., 2002a). According to Grigg et al. (2001b), single locus RFLP analysis of 12 isolates collected from ocular toxoplasmosis cases would have failed to detect 5 recombinant strains and mis-identification of a recombinant strain is likely even with analysis at 3 loci (Grigg et al., 2001b). A multi-locus typing method is essential to understand the true genetic diversity of *T. gondii*, to discover relationships between genotype and human disease (phenotype) and to identify genetic factors which affect virulence.

2.1.2.4 Multi-locus sequencing typing (MLST)

The advent of high throughput sequencing technologies gives us the potential to look at total genetic diversity as defined by application of direct sequencing of genomic regions and discovery of single nucleotide polymorphism (SNP). This sequence-based method is considered to be the best approach for detection of polymorphism, especially for new or non-classified strains (Sibley et al., 2009).

It is essential to determine and understand the processes that control transmission of infectious pathogens, and also to detect and distinguish between different strains of these pathogens. In addition, it is necessary to identify pathogen strains accurately for further epidemiological surveillance and later for the planning of strategies for public health control. To clarify pathogen epidemiology in addition to identification of pathogen isolates, the nucleotide sequence-based technique of multi-locus sequence typing (MLST) was introduced in bacteria (*Neisseria meningitides*) (Maiden et al., 1998). MLST was able to clearly differentiate isolates from different infectious agents by using DNA sequences of usually 7 housekeeping genes - constitutive genes which are essential for the maintenance of basic cellular functions. Thus, MLST provided high-resolution genealogical data.

In the 1980's, the first MLST studies were established on bacterial population structure, and these were essential to the development of the technique (Feil et al., 1999). It was revealed through these studies that genetic exchange through recombination played a dominant role in

the evolution of the majority of prokaryotes (Maiden, 2006). The predominant (clonal model) of bacterial populations was therefore changed to a broader view of partially clonal models. The MLST technique was the major driving force in investigating the extent of the genetic structure of bacterial populations which then quickly developed to be the cornerstone procedure for molecular typing of different pathogenic organisms (Maiden, 2006).

In applying MLST, only a small part of the entire genome of the organism is used. Normally, probes are based on 6 to 7 housekeeping genes and these are suggested to be a representative sample of diversity of the whole genome (Didelot and Maiden, 2010). Protein-encoding housekeeping genes are considered to be the most reliable genetic markers, selected because they evolve slowly and provide high reliable data for both phylogeny analysis and accurate typing.

Although the levels of genetic polymorphism in housekeeping genes are sufficiently high for assessment of strain relationships and population structure (Maiden, 2006), arguments remain about how much genetic diversity, in addition to the association between phylogenetic resolution and gene function (Cooper and Feil, 2006), is essential to accurately understand the evolutionary relationships between species. For instance, studies by Kuhn et al. (2006), Robinson et al. (2005) and Cooper and Feil (2006) studying the pathogen *Staphylococcus aureus* revealed that the addition of rapidly evolving genes did not interfere with the accuracy of conclusions of evolutionary parameters. In the same studies, it was also shown that standard MLST genes provided the lowest phylogenetic resolution. Thus, it was suggested that selection of additional loci for further analysis should be based on nucleotide diversity rather than gene function, at least at the intra-species level (Cooper and Feil, 2006). The conclusion drawn from these studies is that the inclusion of more fast evolving genes might be more valuable than including more standard MLST genes when greater resolution is required (Maiden, 2006).

In *Toxoplasma gondii*, single nucleotide polymorphisms (SNPs) are the most comprehensive group of genetic markers, identified through sequencing of single copy genes. SNPs have been located in the coding regions of different genes for major antigens found on the parasite's surface, or from key internal organelles in rhoptries or dense granules. These surface antigens represent a family of proteins that includes all the surface antigens in *T. gondii*, such as SAG1, 2 and 3, also known as SAG1- related sequences (SRS) genes, which have a significant role in the parasite's attachment and invasion of the host cell. SRS genes

are located in tandem arrays on most Toxoplasma chromosomes (Jung et al., 2004). For example, the allelic variation of SAG1 and 2 has been used widely to differentiate *T. gondii* strains, in addition to distinguishing between mouse-virulent and avirulent strains (Howe and Sibley, 1995; Howe et al., 1996). The other important families of secretory proteins in *T. gondii* are rhoptry (ROPs) and dense granules (GRA) proteins, which have important roles in invasion and intracellular survival of the parasite. For instance, ROP18 is a highly polymorphic protein which has been revealed to play a major role in mouse-virulence, as detected by bioinformatics analysis (Saeij et al., 2006). In addition, the high polymorphic and variable characteristics of the GRA6 gene sequence have been applied widely to differentiate *T. gondii* strains (Fazaeli et al., 2000). In addition, SNPs have been revealed in genes encoding enzymes, such as dihydrofolate reductase and nucleoside triphosphatase (Asai et al., 1995; Aspinall et al., 2002; Binas and Johnson, 1998), in genes encoding structural proteins, such as tubulin or actin (Lehmann et al., 2000), and in genes of unknown functions, such as L328, C19, and B1 (Grigg et al., 2001a; Sibley and Boothroyd, 1992).

2.1.2.5 Whole genome direct sequencing

The complete genome sequences from the reference strains, GT1, ME49 and VEG strains (representing types I, II and III respectively) were generated by the Institute for Genomic Research (TIGR). This research was accompanied by further work on chromosomes 1a and 1b from the type I RH strain, the two smallest chromosomes, which were completed at the Wellcome Trust Sanger Centre (Khan et al., 2006a). This is rapidly expanding and there is now genome data for over 65 strains (http://www.toxodb.org). This information means that primers could be designed for any region of the genome. PCR sequencing can then provide detailed information about the level of polymorphism in specific genes or regions of chromosomes. This high-density information can show more exact relationships between strains and detect recombination with higher efficiency. A greater density of information across the genome also means that rates of mutation can be compared within different groups of genes as evidence of selection (Bontell et al., 2009).

Recently, 200 new SNP-RFLP markers have been developed to map 300-kb intervals, in addition to 50 existing markers, through the 14 chromosomes of the genome of *T. gondii* (Khan et al., 2005). These SNPs have been detected by alignment and comparison with the reference type I (RH), II (ME49) and III (CTG) strains and have been applied to detect the recombination and genetic variation among the predominant strains of *T. gondii*. Using these

markers, it was confirmed that each locus in the main three lineages of *T. gondii* consists of only two alleles, supporting previous findings (Grigg et al., 2001a; Su et al., 2003).

SNP-RFLP markers have a major role in mapping genes associated with phenotypic traits, such as virulence, in epidemiological and experimental studies.

2.1.3 Major Strain Haplotypes

Although Howe and Sibley (1995) suggested the presence of three main lineages, type I, II and III, of *T. gondii*, phylogenetic studies conducted at that time revealed that Toxoplasma has only two major clonal groups (Ajzenberg et al., 2002a; Lehmann et al., 2000). In fact, group 2 is genetically heterogeneous and type III is considered as a sub-group of group 2 (Ajzenberg et al., 2002a). The observation by Grigg et al. (2001a) that many loci of *T. gondii* isolates display only 2 alleles (biallelic), suggested that the *T. gondii* population had only two different ancestries and that the three predominant lineages only resulted from successful recombination between them.

In newly-explored regions, analysis of the population structure may be difficult for various reasons. Firstly, it is predicted that there will be failure of markers to detect genetic diversity in a new area where those markers have been used to identify polymorphisms in existing populations. Secondly, the problematic requirement for intensive sampling strategies to detect allelic frequencies. Thirdly, the absence of exact definitions of the extent to which lineages should be divergent in order to be considered as unique. Although individual isolates within a clonal lineage reveal high levels of linkage disequilibrium and limited recombination, it is not suggested that they are identical (Tibayrenc et al., 1991; Tibayrenc et al., 1993). This is in agreement with minor differences that occur in common alleles due to occasional mutations. Therefore, considering new isolates as unique, solely because they differ in a certain trait of SNPs, is not sufficient to classify them according to genetic diversity. If all isolates genomes were sequenced, they would all be genetically unique. Thus, suitable criteria for classification of strains are required to permit meaningful comparison.

Methods which are used for allocating the genotypes of new isolates into suitable groups include STRUCTURE analysis (Falush et al., 2003), based on multilocus genotype data, other phylogenetic methods, and network analysis (Templeton et al., 1992). For such analysis, the choice of loci involves both regions that are selectively neutral, for common

ancestry estimation, and regions that are under strong selection (antigens) for estimation of diversity. These comparative analyses have the capability to differentiate between minor variants that result from mutation and to identify new lineages which are unique to new regions, or recombinant genotypes that developed from sexual reproduction. Analysis of these local variations in genetic population structure is highly related to phenotypes like pathogenesis and transmission.

Khan et al. (2007) analyzed *T. gondii* population structure based on phylogenetic reconstruction of 46 strains from Europe, North and South America based on 8 intron sequences of 5 unlinked loci and apicoplast genomes. He found 11 haplogroups representing the major lineages which reflected the geographical distribution of *T. gondii*. Most of the strains from North America and Europe were grouped in 1, 2 and 3 haplogroups, which corresponded to the main three clonal lineages, which were recognized in North America and Europe; however, South American strains were distributed between most of the other haplogroups, reflecting the high level of diversity of these strains.

It was suggested that entire of 11 haplogroups might be reconstructed from only four ancestral lineages through very few genetic crosses (Khan et al., 2007). This is similar to the previous prediction that the main three North American lineages arose from only a few genetic crosses (Boyle et al., 2006). Therefore, the population structure of *T. gondii* had been shaped through infrequent random recombination over a long period of time.

The population structure of *T. gondii* was re-evaluated through using sequence-based phylogenetic and population analysis (Khan et al., 2011). This study was based on 5 introns within 5 unlinked loci in addition to 3 antigen encoding genes (SAG1, GRA6 and 7) on expanded sets of isolates of 66 strains from Europe, North and South America. It revealed clustering of 12 haplogroups within *T. gondii* strains. Additionally, a fourth clonal lineage was reported in North America, referred to as haplogroup 12. This study confirmed the previous findings, showing a significant geographical separation between European and North American strains compared with the higher diversity South American strains.

A comprehensive study by Su et al. (2012) set out to genotype 138 isolates of *T. gondii* collected from around the world by application of three different sets of polymorphic DNA markers, including RFLP, microsatellites, and sequencing of introns from housekeeping genes, and 30 different loci distributed across all 14 nuclear chromosomes and plastid

genome. These markers included 11 RFLP genetic markers distributed across 8 of 14 chromosomes in addition to one marker for the apicoplast, 15 microsatellite markers allocated among 10 of 14 chromosomes and 4 introns of 3 different genes. The marked genetic diversity of 138 unique genotypes was arranged, by using clustering methods, into 15 different haplogroups that joined together to form six major clades (Figure 2.1). This study confirmed the predominance of clonal lineages in Northern Hemisphere regions, while in regions of South America it showed higher genetic diversity.

It is obvious that, based on the above studies, we have a good understanding of global diversity; however, it is still difficult to apply this understanding to a more local or regional context. There is evidence that isolates from one geographical area, which might be predicted to be very similar, actually have very different phenotype. In a study by Bontell et al. (2009), 8 *T. gondii* strains that had been collected from Ugandan chickens were genotyped as type II strains (TgCkUg1, 3, 5, 7, 8 and 9), type III (TgCkUg6) and recombinant strains (TgCkUg2) based on five PCR-RFLP markers (SAG1, SAG2, SAG3, GRA6 and BTUB). Additionally, *in vitro* and *in vivo* growth criteria were assessed at the time of isolation. These Ugandan strains were avirulent in mice but while none of the infected mice died or manifested disease symptoms from infection, the parasite burden in the brain, heart, lung and muscle of mice showed significant variations among isolates. The parasite density in the brain of each infected mouse was 15 times the density in quadriceps muscle, 100 times the quantity in heart muscle and 1000 times the amount in lung tissues. In addition, there was shown to be an association between the parasite genotype and tissue burden in mice, where the type III (TgCkUg6) strain produced greater parasite densities compared with the type II strains (TgCkUg1, 3, 5, 7, 8 and 9), while the recombinant strain (TgCkUg2) had an intermediate phenotype. However, this relationship between parasite genotype and densities in mice tissues did not correlate with the growth rate in tissue culture. In this assay, 3 of type II (TgCkUg1, 3 and 8) strains had higher growth rates compared with the type III (TgCkUg6) and recombinant (TgCkUg2) strain with intermediate rates, while the remaining type II strains (TgCkUg5, 7 and 9) had low growth rates.

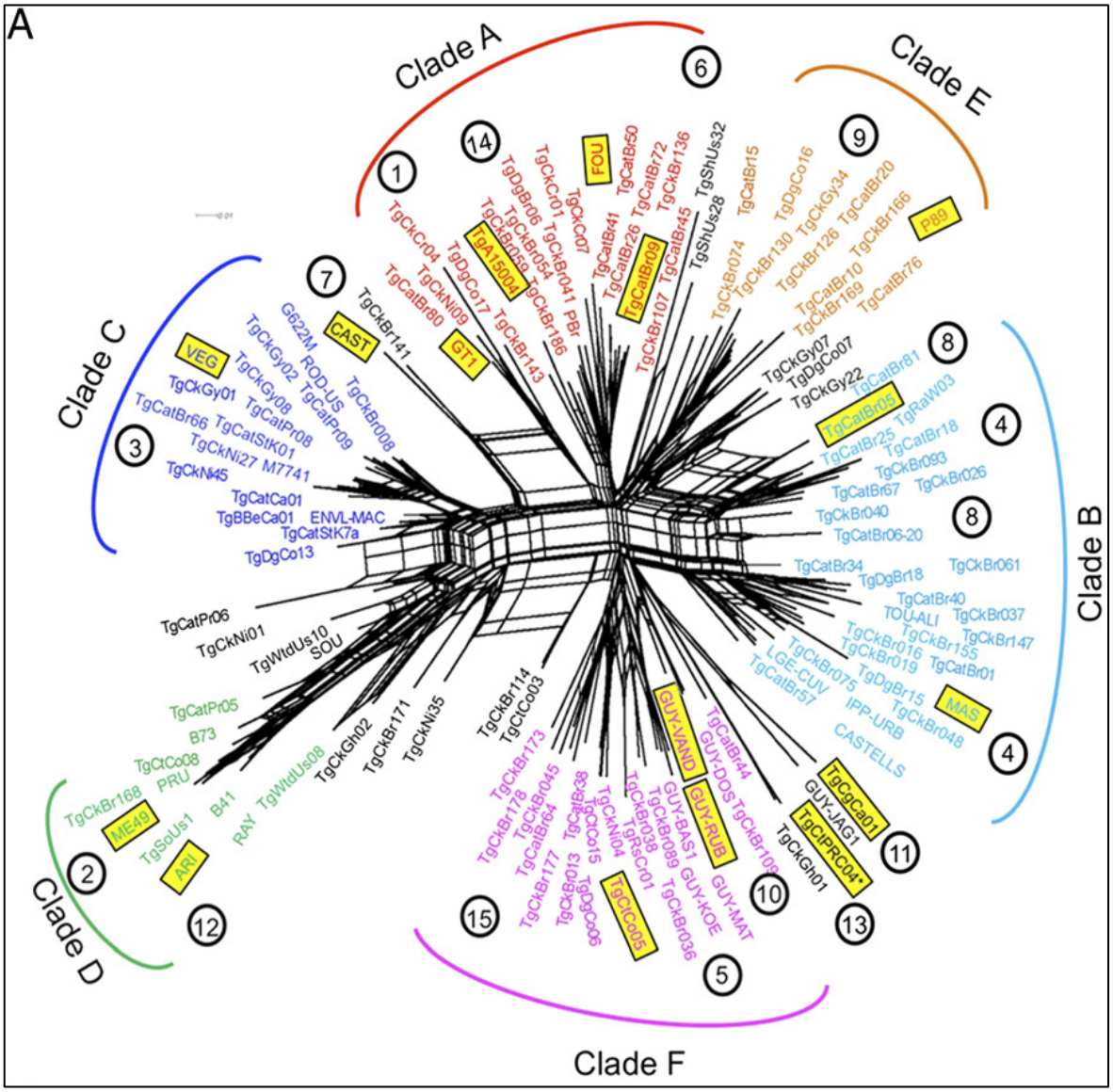

Figure 2.1: Population genetic structure of *Toxoplasma gondii* (Su et al., 2012)

2.1.4 Aims

In this study, TgCkUg1, 3, 5, 6, 7, 8 and 9 strains were subjected to more intensive sequencing with a panel of the most frequently used PCR RFLP markers, which have previously been used to detect variation within different *Toxoplasma gondii* strains collected from various regions in the world (Su et al., 2010; Su et al., 2012);to toprovide more information through sequencing these loci , to seek evidence of local allelic variation and to test whether differences in the phenotype of strains related to genetic variation at key loci.

2.2 METHODS

2.2.1 Parasite Strains

Eight *T. gondii* strains collected by Bontell et al. (2009) from Ugandan chickens were used in this study. The chickens were obtained from two different regions in and around the capital city of Uganda, Kampala. These regions were Gayaza, in the north of Kampala, (TgCkUg1, 2 & 3) and Mulago, near to the city center of the capital city, (TgCkUg5, 6, 7, 8 & 9) (Lindstroem et al., 2008). These strains were designated as TgCkUg1, 2, 3, 5, 6, 7, 8 and 9 (Table 2.5) *in vitro* and *in vivo* growth criteria were assessed at the time of isolation. One isolate TgCkUg2 had been subjected to whole genome sequencing while in other isolates a restricted number of loci were sequenced to investigate their genotype (Table 2.5). In this study, TgCkUg1, 3, 5, 6, 7, 8 and 9 strains were subjected to deeper sequencing.

Table 2.5: Genotype and phenotype analysis of the Ugandan strains (Bontell et al., 2009)

No.	**Isolate**	**Genotypes** Based on sequencing of three PCR-RFLP markers (SAG1, SAG3 and GRA6)	**Phenotypes**	
			In vitro (Growth rate)	***In vivo*** (Parasite density, tissue burden)
1	TgCkUg1	II	High	Low
2	TgCkUg2	II + III	Intermediate	Intermediate
3	TgCkUg3	II	High	Low
4	TgCkUg5	II	Low	Low
5	TgCkUg6	III	Intermediate	High
6	TgCkUg7	II	Low	Low
7	TgCkUg8	II	High	Low
8	TgCkUg9	II	Low	Low

2.2.2 Cell culture

In order to propagate parasites for genetic analysis they were cultured *in vitro* using human foreskin fibroblast (HFF) cell line. HFF cells were grown in 25 cm^2 flask in 10 ml of culture medium incubated at 37.0 ^{0}C in a 5% CO_2 atmosphere. The culture medium consists of Dulbecco's Modified Eagle's Medium (DMEM) with L- Glutamine supplemented with 10% foetal calf serum (FCS). The cells were sub-cultured when the confluent monolayer was obtained, normally at weekly intervals with split ratio of 1:4. Sub-culturing was achieved by

trypsinisation (Trypsin/EDTA, 0.04% / 0.03%) of the monolayer. 10 ml of phosphate buffered saline (PBS, pH 7.3) was used for monolayer washing. 10 ml of Trypsin/EDTA was then added and left for 2 minutes. 8 ml of Trypsin/EDTA was removed and the flask containing monolayer cells with 2 ml Trypsin/EDTA was incubated at 37.0 ^{0}C in a 5% CO_2 atmosphere for 2-5 minutes. Cells were released by gently tapping the flask; this resulted in detachment of cells from the flask surface. Cells were observed under an inverted microscope until they detached.

2.2.3 Parasite culture

25cm^2 flasks of HFF cells were infected with 10^6 viable tachyzoites of the type III NED strain of *T. gondii*. *T. gondii* cultures were examined using an inverted microscope to monitor the presence of tachyzoites. The cultures were passaged when maximally infected, as revealed by the appearance of plaques in the cell monolayer. Tachyzoites were harvested by release from the monolayer by sharply tapping the side of the flask on the bench. To ensure that the parasite had been released, the monolayer was examined by microscopy. The released tachyzoites in the supernatant were counted by haemocytometer. Parasites were sub-cultured into a new 25 cm^2 flask containing a confluent monolayer of HFF cells with complete growth medium and incubated at 37.0 ^{0}C with 5% CO_2. The remaining parasite suspension was transferred to a centrifuge tube and spun down for 15 min at 2500 r.p.m. The parasite pellets were stored in -20^{0}C for DNA extraction.

2.2.4 DNA extraction

DNA from culture-derived tachyzoites of *T. gondii* isolates was extracted and purified by using the Qiagen DNeasy Blood and Tissue Kit. The pellets were resuspended in 200 µl of PBS followed by the addition of 20 µl of proteinase K (600 mAU/ml solution or 40 mAU/mg protein). After adding 200 µl Lysis Buffer (AL), the mixture was incubated at 56^{0}C for 1 hour to lyse the parasites and release the DNA molecules. After the addition of 200 µl ethanol (96-100%), the mixture was transferred to a DNeasy Mini spin column. The columns were washed twice with Wash Buffers (AW1) and (AW2). The genomic DNA was eluted from the columns by adding 200 µl of Elution Buffer (AE) (10 mM Tris·Cl, 0.5 mM EDTA, pH 9.0) to the spin column and incubating for 1 min at room temperature (15-25^{0}C). Nanodrop was used to measure the concentrations and purity for all samples prior to PCR analysis. The genomic DNA was stored at -20^{0}C until use.

2.2.5 Multi-locus nested PCR

Multi-locus nested PCR analysis of *Toxoplasma gondii* samples (NED type III strain and the Ugandan collection of seven *T. gondii* strains TgCkUg1, 3, 5, 6, 7, 8 and 9) was conducted, using twelve different genetic markers distributed across eight chromosomes and the apicoplast genome (SAG1, 5'-SAG2, 3'-SAG2, Alt.SAG2, GRA6, L358, BTUB, SAG3, C22-8, C29-2, PK1 and Apico) (Table 2.6).

Table 2.6: Multi-locus nested PCR external and internal primers (Su et al., 2010)

No.	Markers	Chromosomes	Nested PCR (bp)	Multiplex PCR primers (external primers)	Nested PCR primers (internal primers)
1	**SAG1**	VIII	390	F: GTTCTAACCACGCACCCTGAG R: AAGAGTGGGAGGCTCTGTGA	F: CAATGTGCACCTGTAGGAAGC R: GTGGTTCTCCGTCGGTGTGAG
2	**5`-SAG2**	VIII	242	Not needed. (The DNA fragment for 5`-SAG2 is covered by the external primers of alt. SAG2)	F: GAAATGTTTCAGGTTGCTGC R: GCAAGAGCGAACTTGAACAC
3	**3`-SAG2**	VIII	222	F: TCTGTTCTCCGAAGTGACTCC R: TCAAAGCGTGCATTATCGC	F: ATTCTCATGCCTCCGCTTC R: AACGTTTCACGAAGGCACAC
4	**alt. SAG2**	VIII	546	F: GGAACGCGAACAATGAGTTT R: GCACTGTTGTCCAGGGTTTT	F: ACCCATCTGCGAAGAAAACG R: ATTTCGACCAGCGGGAGCAC
5	**SAG3**	XII	225	F: CAACTCTCACCATTCCACCC R: GCGCGTTGTTAGACAAGACA	F: TCTTGTCGGGTGTTCACTCA R: CACAAGGAGACCGAGAAGGA
6	**BTUB**	IX	411	F: TCCAAAATGAGAGAAATCGT R: AAATTGAAATGACGGAAGAA	F: GAGGTCATCTCGGACGAACA R: TTGTAGGAACACCCGGACGC
7	**GRA6**	X	344	F: ATTTGTGTTTCCGAGCAGGT R: GCACCTTCGCTTGTGGTT	F: TTTCCGAGCAGGTGACCT R: TCGCCGAAGAGTTGACATAG
8	**C22-8**	Ib	521	F: TGATGCATCCATGCGTTTAT R: CCTCCACTTCTTCGGTCTCA	F: TCTCTCTACGTGGACGCC R:AGGTGCTTGGATATTCGC
9	**C29-2**	III	446	F: ACCCACTGAGCGAAAAGAAA R: AGGGTCTCTTGCGCATACAT	F: AGTTCTGCAGAGTGTCGC R:TGTCTAGGAAAGAGGCGC
10	**L358**	V	418	F: TCTCTCGACTTCGCCTCTTC R: GCAATTTCCTCGAAGACAGG	F: AGGAGGCGTAGCGCAAGT R: CCCTCTGGCTGCAGTGCT
11	**PK1**	VI	903	F: GAAAGCTGTCCACCCTGAAA R: AGAAAGCTCCGTGCAGTGAT	F: CGCAAAGGGAGACAATCAGT R: TCATCGCTGAATCTCATTGC
12	**Apico**	Plastid	640	F:TGGTTTTAACCCTAGATTGTGG R:AAACGGAATTAATGAGATTTGAA	F:GCAAATTCTTGAATTCTCAGTT R: GGGATTCGAACCCTTGATA

The first round of PCR reaction was achieved in a total volume of 25 µl which contained 2 mM $MgCl_2$, 1x PCR buffer, 0.15 µM each of the external forward and reverse primers, 200 µM each of the dNTPs, 1 unit of DNA polymerase and 1.5 µl of DNA samples. This reaction mixture was first heated at 95 ^{0}C for 4 minutes, followed by 30 cycles of 94 ^{0}C for 30 seconds, 55 ^{0}C for 1 minute and 72 ^{0}C for 2 minutes. For dilution of PCR amplified products (1:1), 25 µl of nuclease free water was added to the mixture.

Nested PCR was achieved in a total volume of 25 µl containing 2 mM $MgCl_2$, 1x PCR buffer, 0.30 µM each of the internal forward and reverse primers, 200 µM each of the dNTPs, 1 unit of DNA polymerase and 1.5 µl of diluted PCR products. This reaction mixture was first treated at 95 ^{0}C for 4 minutes, followed by 35 cycles of 94 ^{0}C for 30 seconds, 60 ^{0}C for 1 minute and 72 ^{0}C for 1.5 minutes (for the Apico marker, the annealing temperature is at

58 ^{0}C instead of 60 ^{0}C). Calculations have been made to determine the amounts of each component of both rounds of nested PCR.

The PCR products were examined by electrophoresis of 5 µl of PCR products in 1.5% agarose gel containing 1.5 g agarose gel, 100 ml 1X TBE buffer and 100 µl Gel Red and visualized under UV light.

2.2.6 Sequencing

PCR products for all eleven genetic markers (SAG1, 5'-SAG2, 3'-SAG2, Alt.SAG2, GRA6, L358, BTUB, C22-8, C29-2, PK1 and Apico) of *Toxoplasma gondii* samples (NED type III strain and the Ugandan collection of seven *T. gondii* strains, TgCkUg1, 3, 5, 6, 7, 8 and 9) were purified prior to DNA sequencing by using the Qiagen DNeasy Blood and Tissue Kit. Nanodrop was used to measure the concentrations and purities for all samples prior to being sent to Beckman Coulter Genomics for sequencing (forward and reverse). The sequence results were viewed by using FinchTV 1.5.0 (Figure 2.4). Results were compared against published archetypal strain sequences in GenBank using The Basic Local Alignment Search Tool (BLAST).

2.2.7 Phylogenetic analysis

The sequences of PCR- amplified regions were aligned with data from the reference strains of clonal type I, II and III strains (GT1, ME49, and VEG) respectively available at (http://www.toxodb.org/toxo/) by using ClustalW. Then the aligned sequences were incorporated into Molecular Evolutionary Genetic Analysis (MEGA) version 5.1 (http://www.megasoftware.net) for identification of single nucleotide polymorphism (SNPs) and construction of Neighbor-Joining phylogenetic trees. These sequences will be deposited in USIR (University of Salford Institutional Repository) and Genbank.

2.3 RESULTS

The main focus of this research project is to investigate the level of genetic variation among *Toxoplasma gondii* isolates in a cluster of strains obtained from a single area at a single time point and to relate this to their phenotype. The starting point for the research was a collection of isolates from Africa, TgCkUg1, 3, 5, 6, 7, 8 and 9, (Table 2.5), which appear to show homology to the main clonal lineages type II and type III. The aims are to seek evidence of local allelic variation and to test whether differences in the phenotype of strains relate to genetic variation at key loci by using multi-locus typing and then subsequently using deeper sequencing. In order to facilitate this analysis I have allocated *in vivo* and *in vitro* pathogenesis scores based on the data from Bontell et al. (2009).

2.3.1 Multi-locus typing of all loci using NED type III strain for optimization

In initial studies PCR optimization was carried out by applying eleven different genetic markers distributed across eight chromosomes and the apicoplast genome (SAG1, 5'-SAG2, 3'-SAG2, Alt.SAG2, GRA6, L358, BTUB, SAG3, C22-8, C29-2, PK1 and Apico) to the NED type III strain of *Toxoplasma gondii*. The loci were distributed across eight chromosomes**.** For all loci, except SAG3, multi-locus nested PCR procedure was conducted as described by Su et al. (2010) except that the amounts of DNA sample in the first round and of diluted PCR products in the second round was 1 µl instead of 1.5 µl. SAG3 primers failed to produce positive results even after changing the annealing temperature of both PCR rounds (temperature gradient) and modifying the concentration of $MgCl_2$.

All loci were successfully amplified with clear DNA bands visualized by gel images and met the predicted band sizes (Table 2.5), except for SAG3. Figure 2.2 shows examples of gel images of nested PCR products for the NED strain of *Toxoplasma gondii* by using 3'-SAG2, 5'-SAG2, GRA6 and C29-2 genetic markers.

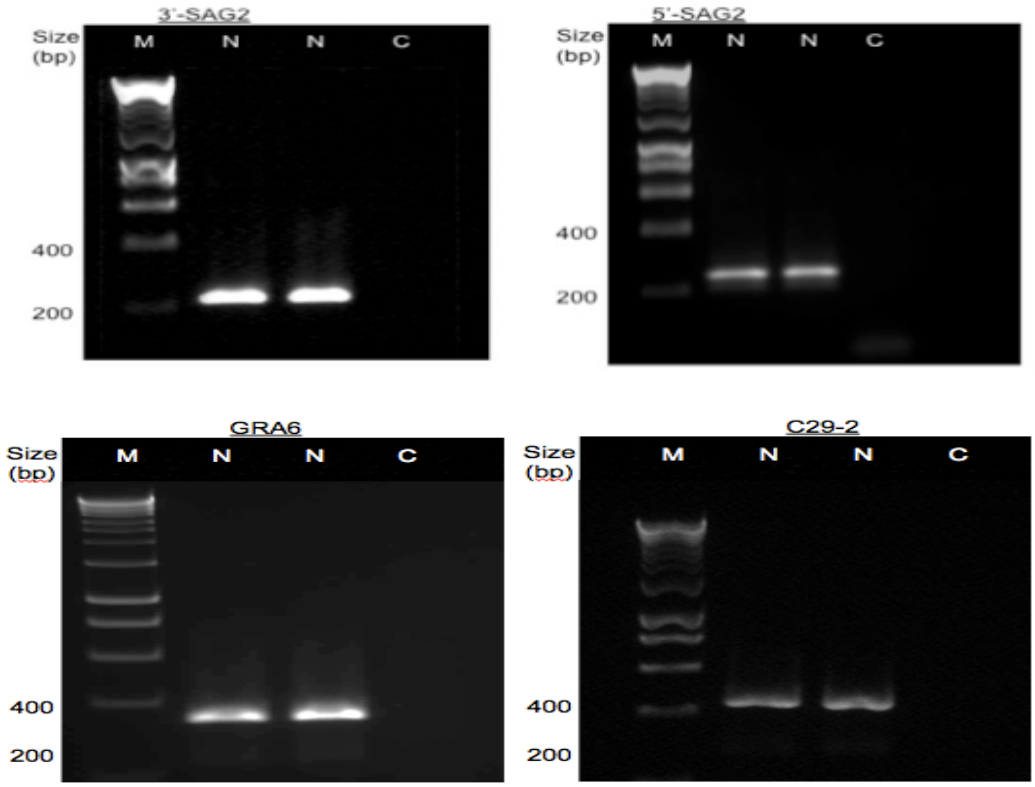

Figure 2.2: Gel images of nested PCR products for NED strain of *Toxoplasma gondii* using 3'-SAG2, 5'-SAG2, GRA6 and C29-2 genetic markers
(M: molecular weight DNA ladder, N: NED strain, C: negative control)

2.3.2 Multi-locus typing of Ugandan strains

Having optimized conditions, multi-locus nested PCR was applied to the Ugandan collection of seven *T. gondii* strains (Table 2.5). Twelve different genetic markers were distributed across eight chromosomes and the apicoplast genome (SAG1, 5'-SAG2, 3'-SAG2, Alt.SAG2, GRA6, L358, BTUB, SAG3, C22-8, C29-2, PK1 and Apico). Table 2.7 shows the measurement of concentrations and confirmation of purity (by values of >= 1.8 for the Nanodrop 260/280 ratio) of the purified DNA for all samples, by using Nanodrop, prior to start PCR analysis.

Table 2.7: The measuring concentrations and confirmation of purity of the purified DNA for all Ugandan samples by using Nanodrop

Samples (DNA)	Nucleic acid concentration (ng/ µl)	260/280
TgCkUg1	13.0	1.96
TgCkUg3	27.8	1.89
TgCkUg5	14.8	1.94
TgCkUg6	21.1	1.85
TgCkUg7	26.0	1.89
TgCkUg8	21.2	2.01
TgCkUg9	19.5	1.92
NED	22.8	1.95

As shown in Figure 2.3 below, all loci except for SAG3 were successfully amplified with clear DNA bands visualized by UV as gel images.

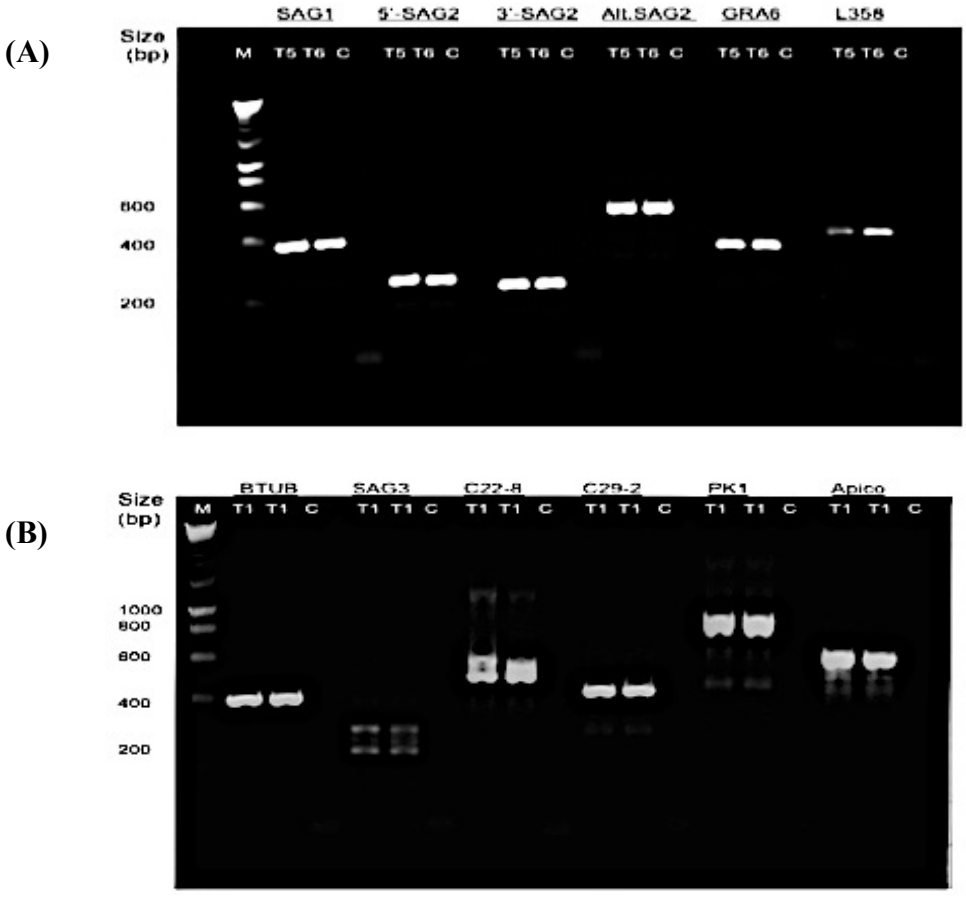

Figure 2.3: Gel images of nested PCR products for African isolates of *Toxoplasma gondii* gained by using 12 different genetic markers
(A) TgCkUg5 and TgCkUg6 using SAG1, 5'-SAG2, 3'-SAG2, Alt.SAG2, GRA6 and L358 (B) TgCkUg1 using BTUB, SAG3, C22-8, C29-2, PK1 and Apico. (M: molecular weight DNA ladder, T5: TgCkUg5 strain, T6: TgCkUg6 strain, T1: TgCkUg1 strain, C: negative control)

Taking the data overall, there was a variation in the frequency of successful nested PCR amplifications between the different loci (Table 2.8). SAG1, PK1 and Apico were the most successful markers, with 100% positive PCR results. By contrast, SAG3 demonstrated totally negative results with failure in PCR amplification of all nested PCR experiments. Most of the other markers showed an acceptable success rates at 86% while the BTUB marker was at 71%.

Table 2.8: Success rates of all loci via multi-locus nested PCR procedure

Locus	No. of PCR	No. of +ve results	No. of –ve results	% of +ve results
SAG1	7	7	0	100
5'-SAG2	7	6	1	86
3'-SAG2	7	6	1	86
Alt. SAG2	7	6	1	86
SAG3	7	0	7	0
BTUB	7	5	2	71
GRA6	7	6	1	86
C22-8	7	6	1	86
C29-2	7	6	1	86
L358	7	6	1	86
PK1	7	7	0	100
Apico	7	7	0	100

2.3.3 Sequence, alignment and SNPs

Sequence data (both forward and reverse) was generated from all isolates, for eleven genetic markers distributed across seven chromosomes and the apicoplast genome.

The forward sequences were successfully generated for eleven genetic markers from all isolates, except from TgCkUg1 at PK1 locus, while generation of the reverse sequences failed from TgCkUg1 and TgCkUg6 at C29-2 locus, and TgCkUg7 at L358 locus. The measurement of concentrations and purity for all samples was applied by using Nanodrop, prior to being sent for sequencing (forward and reverse) which was carried out by Beckman Coulter Genomics. The purity of DNA is indicated by the reading 260/280 >= 1.8. The sequence results were viewed by using FinchTV 1.5.0. (Figure 2.4).

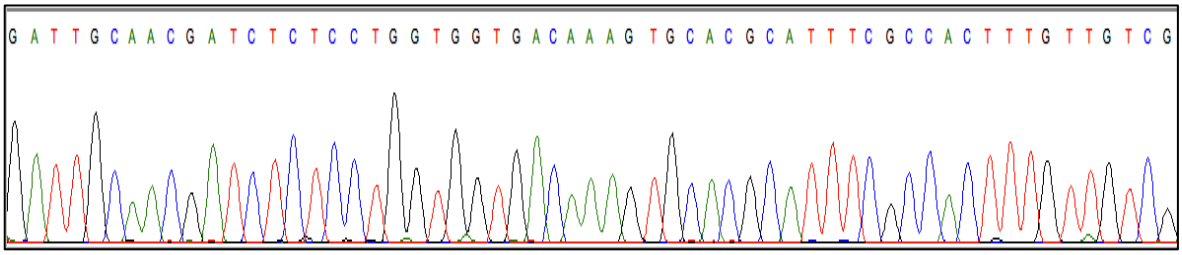

Figure 2.4: Example of viewing DNA chromatograph by using FinchTV 1.5.0

The resulting data were aligned with sequences from the reference strains of clonal type I, II and III strains (GT1, ME49, and VEG respectively) available at (http://www.toxodb.org/toxo/) by using ClustalW. The aligned sequences were then incorporated into the Molecular Evolutionary Genetic Analysis software package (MEGA version 5.1) for identification of single nucleotide polymorphisms (SNPs) and construction of Neighbor-Joining phylogenetic trees. Figure 2.5 shows the variable sites at 8 genetic markers (C22-8, C29-2, L358, PK1, SAG1, SAG2, BTUB and GRA6) by direct PCR-DNA sequencing of *T. gondii* isolates of NED and Ugandan strains (TgCkUg1, 3, 5, 6, 7, 8 and 9) compared with the reference strains of clonal type I, II and III (GT1, ME49, and VEG respectively).

Markers	C22-8 -Chr Ib					C29-2 -Chr III				L358-Chr V			PK1- Chr.VI						
Nucleotides	1855089	1855119	1855191	1855399	Allele	663058	663168	663202	Allele	2241588	2241636	Allele	2682728	2682703	2682616	2682418	2682272	2682265	Allele
Consensus	A	A	T	G		C	T	G		G	C		C	G	A	G	G	A	
GT1	.	.	C	A	I	.	G	.	I	.	G	I	.	.	.	.	.	.	I
ME49	.	.	.	.	II	T	.	.	II	A	.	II	.	.	.	T	C	.	II
VEG	T	G	.	.	III	.	.	C	III	.	.	III	T	A	T	.	.	.	III
NED	T	G	.	.	III	.	.	C	III	-	-	III	T	A	T	.	.	.	III
TgCkUg1	.	.	.	.	II	T	.	.	II	A	.	II	.	.	.	T	C	.	II
TgCkUg3	.	.	.	.	II	T	.	.	II	A	.	II	.	.	.	T	C	.	II
TgCkUg5	T	G	.	.	III	.	.	C	III	.	.	III	T	A	T	.	.	.	III
TgCkUg6	.	.	.	.	II	T	.	.	II	A	.	II	.	.	.	T	C	.	II
TgCkUg7	T	G	.	.	III	.	.	C	III	A	.	II	.	.	.	T	C	.	II
TgCkUg8	.	.	.	.	II	T	.	.	II	A	.	II	.	.	.	T	C	.	II
TgCkUg9	.	.	.	.	II	T	.	.	II	A	.	II	.	.	.	T	C	G	u-1

Figure 2.5: SNPs pattern at 8 genetic markers (C22-8, C29-2, L358, PK1, SAG1, SAG2, BTUB and GRA6) by direct PCR-DNA sequencing of *Toxoplasma gondii* isolates of NED and Ugandan strains (TgCkUg1, 3, 5, 6, 7, 8 and 9) compared with the reference strains of clonal type I, II and III (GT1, ME49, and VEG respectively)
Green SNPs pattern represent a type II background (identical to type II reference - ME49), blue SNPs pattern represent a type III background (identical to type III reference – VEG), and red SNPs pattern represent novel SNPs (different from both references). Similarities to type II reference sequence (TgME49) are indicated by (.).
Numerical positions refer to sites in the published sequences of (ME49) reference strain (GenBank accession no. XM_002368164, JX045474, AF239285, JX045468 for SAG1, SAG2, GRA6, and BTUB respectively). For **C22-8, C29-2, L358 and PK1,** numerical positions refer to sites in the published sequences of (ME49) reference strain chromosomes from http://www.toxodb.org (TGME49_chrIb, TGME49_chrIII, TGME49_chrV and TGME49_chrVI respectively).

Markers	SAG1-ChrVIII			SAG2-Chr VIII					BTUB-Chr IX						GRA6-Chr X					
Nucleotides	46	134	Allele	138	278	289	298	Allele	117	182	219	222	382	Allele	41	71	106	162	171	Allele
Consensus	G	C		T	C	C	G		C	C	C	G	T		C	T	C	G	A	
GT1	A	T	I	.	.	.	.	I	.	.	.	.	.	I	.	G	.	.	.	I
ME49	.	.	II	G	A	G	C	II	G	G	G	C	C	II	T	.	.	.	G	II
VEG	.	.	III	.	.	.	.	III	.	.	.	.	.	III	.	.	T	A	.	III
NED	.	.	II/III	.	.	.	.	I/III	.	.	.	.	C	u-1	.	.	T	A	.	III
TgCkUg1	.	.	II/III	G	A	G	C	II	G	G	G	C	C	II	T	.	.	.	G	II
TgCkUg3	.	.	II/III	G	A	G	C	II	G	G	G	C	C	II	T	.	.	.	G	II
TgCkUg5	.	.	II/III	.	.	.	.	I/III	.	.	.	.	C	u-1	T	.	.	.	G	II
TgCkUg6	.	.	II/III	G	A	G	C	II	G	G	G	C	C	II	.	.	T	A	.	III
TgCkUg7	.	.	II/III	.	.	.	.	I/III	G	G	G	C	C	II	T	.	.	.	G	II
TgCkUg8	.	.	II/III	G	A	G	C	II	G	G	G	C	C	II	T	.	.	.	G	II
TgCkUg9	.	.	II/III	G	A	G	C	II	G	G	G	C	C	II	T	.	.	.	G	II

Figure 2.5: continue

The panel of markers used for global strain analysis was based on a mix of loci that included neutral markers introns from genes that encode housekeeping function (BTUB), coding sequence from surface and secretory proteins, (SAG1, SAG2, GRA6, PK1) and from genes with unknown functions (L358, C22-8 and C29-2). Greater variation between the neutrally evolving intron was expected than in the coding regions of functional genes. However, genes, which encode major antigens, are under selective pressure and might be predicted to have higher variation (Miller et al., 2004; Pena et al., 2008). Table 2.9 shows a variation in estimated SNPs/kb rates among these genetic markers.

Table 2.9: SNP/kb rates estimation in each locus compared with the reference strains of clonal type I, II and III (GT1, ME49, and VEG respectively)

Locus	Number of SNPs	Estimated SNPs/kb	Size (bp)
SAG1	2	5.1	390
SAG2 (5'-SAG2, 3'-SAG2 and Alt.SAG2)	4	4.0	1010 (242 + 222 +546)
GRA6	5	14.5	344
BTUB	5	12.2	411
L358	2	4.8	418
C22-8	4	7.7	521
C29-2	3	6.7	446
PK1	6	6.6	903
Apico	3	4.7	640
TOTAL	**34**	**66.3**	**5083**

This shows that the GRA6 gene sequence had the highest estimated SNPs/kb rate at 14.5 SNPs/kb, followed by BTUB at 12.2 SNPs/kb, while SAG2 genetic marker had the lowest estimated SNPs/kb rate at 4.0 SNPs/kb. Loci encoding surface antigens (SAG1), secretory proteins (PK1), genes with unknown functions (L358, C22-8 and C29-2) and the apicoplast markers had rates estimated between 7.7 SNPs/kb and 4.7 SNPs/kb.

It was then possible to compare all the sequences of all genetic markers within the type II like Ugandan strains with the type II reference strain ME49. The data are shown in Table 2.10. It appears from this comparison that three isolates TgCkUg1, 3 and 8 are identical to ME49 reference strain. Among the other strains, the highest number of SNPs was shown in TgCkUg5 at 19 SNPs distributed across 7 loci, while TgCkUg9 had the lowest SNPs rate with only one SNP within the PK1 locus.

Table 2.10: Number of SNPs in each locus of the seven Ugandan strains compared with the reference strain of clonal type II (ME49)

Locus	SNPs compared to type II reference strain (ME49)						
	TgCkUg1	**TgCkUg3**	**TgCkUg5**	**TgCkUg6**	**TgCkUg7**	**TgCkUg8**	**TgCkUg9**
SAG1	0	0	0	0	0	0	0
SAG2	0	0	4	0	4	0	0
GRA6	0	0	0	4	0	0	0
BTUB	0	0	4	0	0	0	0
L358	0	0	1	0	0	0	0
C22-8	0	0	2	0	2	0	0
C29-2	0	0	2	0	2	0	0
PK1	0	0	5	0	0	0	1
Apico	0	0	1	2	0	0	0
TOTAL	**0**	**0**	**19**	**6**	**8**	**0**	**1**

Figure 2.6 shows that SNPs distribution is variable within NED and Ugandan strains (TgCkUg1, 3, 5, 6, 7, 8 and 9) over 7 of 14 chromosomes of *T. gondii*. The SAG1 marker is not included in this analysis, as it cannot distinguish between type II and III strains. It is obvious that all SNPs among TgCkUg1, 3, 8 and 9 strains over 7 of 14 chromosomes have only a type II background. In contrast, TgCkUg5, 6, 7 and NED strains have a mixture of type II and III background SNPs in addition to novel SNPs within BTUB (Chr. IX) locus in TgCkUg5 and NED strains and PK1 (Chr. VI) locus in TgCkUg9. TgCkUg5 has type II background SNPs in GRA6 (Chr.X) loci, while SNPs within C22-8 (Chr. Ib), C29-2 (Chr.III), L358 (Chr. V), PK1 (Chr. VI) and SAG2 (Chr. VIII) loci has type III background. In addition, type II background SNPs are located in L358 (Chr. V), PK1 (Chr. VI), BTUB (Chr. IX) and GRA6 (Chr.X) loci of the TgCkUg7 strain, while C22-8 (Chr. Ib), C29-2 (Chr.III) and SAG2 (Chr. VIII) loci have type III background SNPs. It is clear from the pattern of SNPs distribution in TgCkUg5, 6, 7 and NED strains that this is likely to be evidence of recombination between type II and III within these strains, to be added to the previously evidenced recombinant Ugandan strain, TgCkUg2, (Bontell et al., 2009) (Table 2.11 and Figure 2.6).

Table 2.11: Comparison between type II and III background SNPs chromosomes in TgCkUg2 (Bontell et al., 2009) with Ugandan strains that exhibit a mixture of type II and III backgrounds SNPs

Strain	**Type II background SNPs chromosomes**	**Type III background SNPs chromosomes**
TgCkUg2 (Bontell et al., 2009)	II, IV, VI, VIIa, IX and X	Ia, Ib, III, V, VIIb, VIII and XII
TgCkUg5	X	Ib, III, V, VI and VIII
TgCkUg6	Ib, III, V, VI, VIII and IX	X
TgCkUg7	V, VI, IX and X	Ib, III and VIII
NED		Ib, III, V, VI, VIII and X

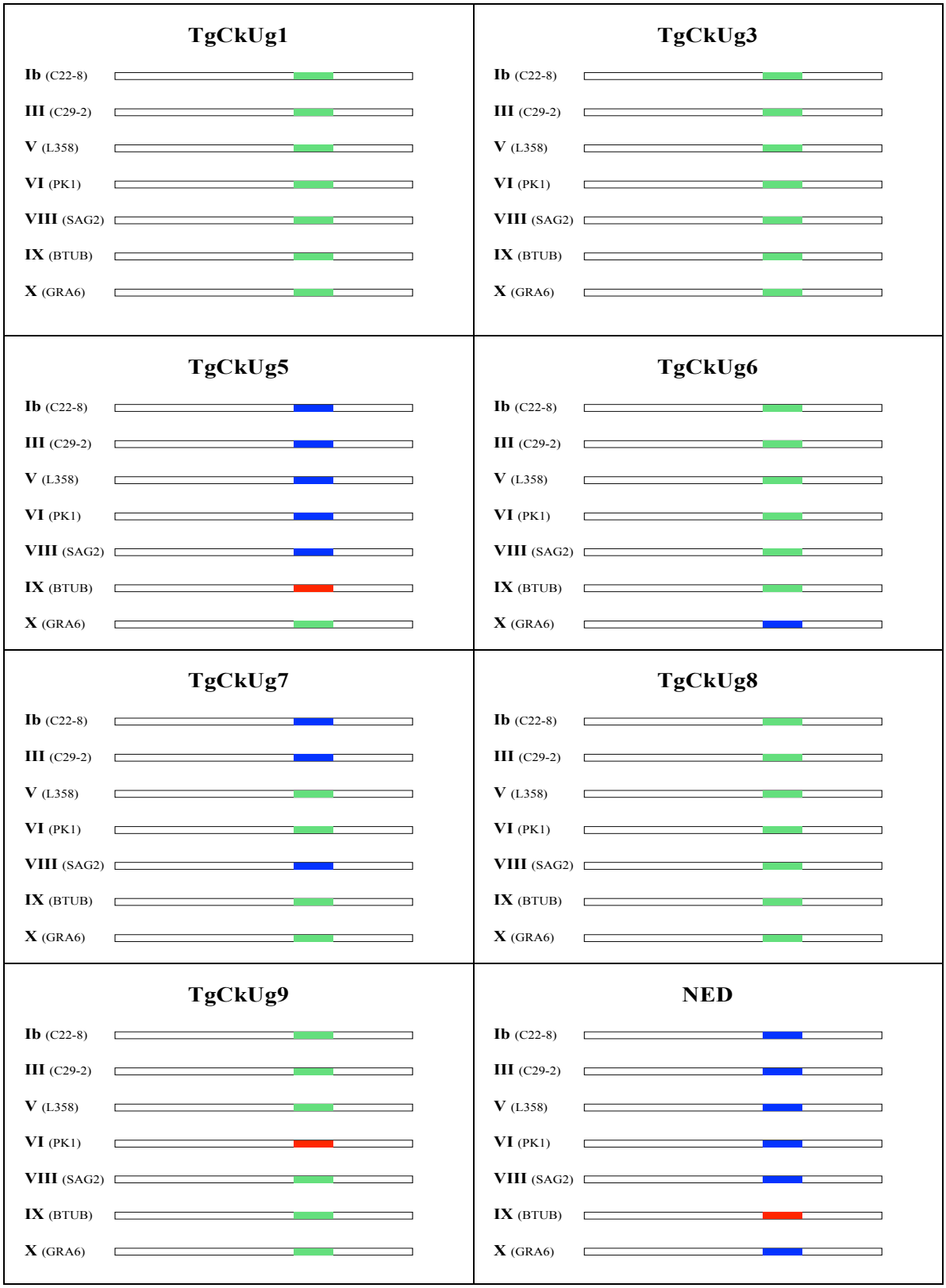

Figure 2.6: SNPs distribution within NED and Ugandan strains (TgCkUg1, 3, 5, 6, 7, 8 and9) over 7 of 14 chromosomes of *Toxoplasma gondii*.
Green SNPs represent a type II background (identical to type II reference - ME49), blue SNPs represent a type III background (identical to type III reference – VEG), and red SNPs represent novel SNPs (different from both references)

2.3.4 Phylogenetic Analysis

To analyze the relationship between strains, sequences of PCR- amplified regions were aligned by using ClustalW then incorporated into molecular evolutionary genetic analysis (MEGA) version 5.1 for construction of Neighbor-Joining phylogenetic trees and identification of single nucleotide polymorphism (SNPs). Trees drawn for single loci illustrate how strains fall into the type strains (Figure 2.7).

For instance, SAG1 locus is shared by type II/III strains, and all African isolates are clustered with the reference strains of clonal type II and III (ME49, and VEG) respectively. All type II Ugandan strains, TgCkUg1, 3, 5, 7, 8 and 9, are clustered with the reference strains of clonal type II at GRA6 locus, while type III one, TgCkUg6, is clustered with the reference type III strain and type III NED strain. In addition, data from an individual locus lacked the ability to detect the variations between different strains. For example, in SAG1 analysis for NED and Ugandan strains it was clear that there is no variation between them. Analysis of these isolates by using different genetic markers show their relations to the reference strains of clonal type II and III (ME49, and VEG). TgCkUg6 strain revealed a strong relation with VEG strains (the reference type III strain) at only one locus (GRA6). At SAG2, BTUB, C22-8, C29-2 and L358 loci, TgCkUg6 displayed a similar pattern to the reference strain type II (ME49). However, TgCkUg5 and 7 strains, which were classified as type II strain, showed similarities to the VEG (III) strain at SAG2, C22-8 and C29-2 loci. In addition, TgCkUg5 demonstrated a strong relation to type III reference strain at BTUB and L358 loci.

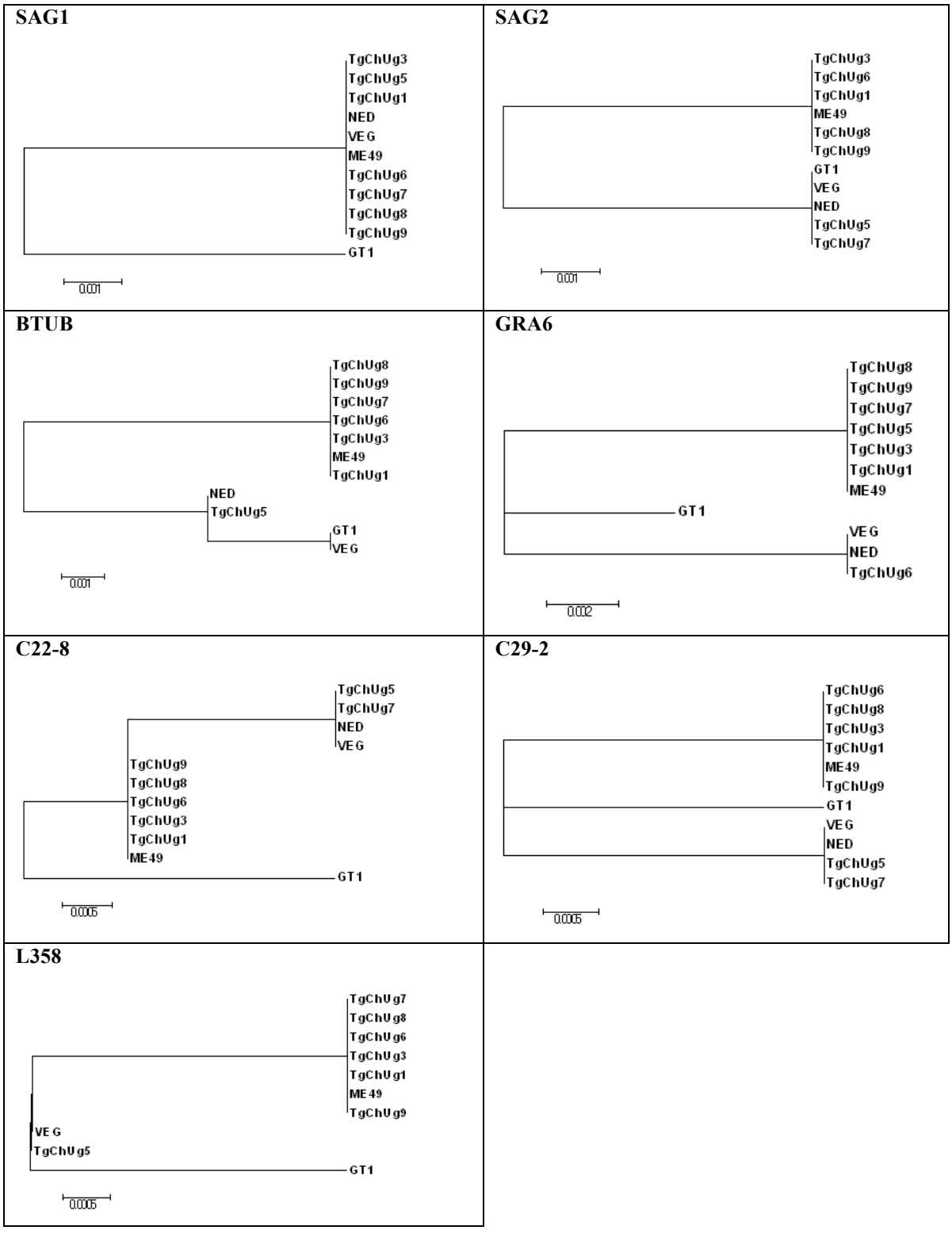

Figure 2.7: Construction of Neighbor-Joining tree using SAG1, SAG2, BTUB, GRA6, C22-8, C29-2 and L358 loci from NED, Ugandan strains (TgCkUg1, 3, 5, 6, 7, 8 and 9) with the reference strains of clonal type I, II and III (GT1, ME49, and VEG respectively)

For deeper analysis, and to detect greater genetic diversity between Ugandan isolates and to be able to segregate them, concatenation of DNA sequences of different loci was applied. Thus, concatenation of DNA sequences of 9 loci (SAG1, SAG2, BTUB, GRA6, L358, C22-8, C29-2, PK1 and Apico) from NED and Ugandan strains was performed to be able to identify a higher level of variation within African isolates (Figure 2.8).

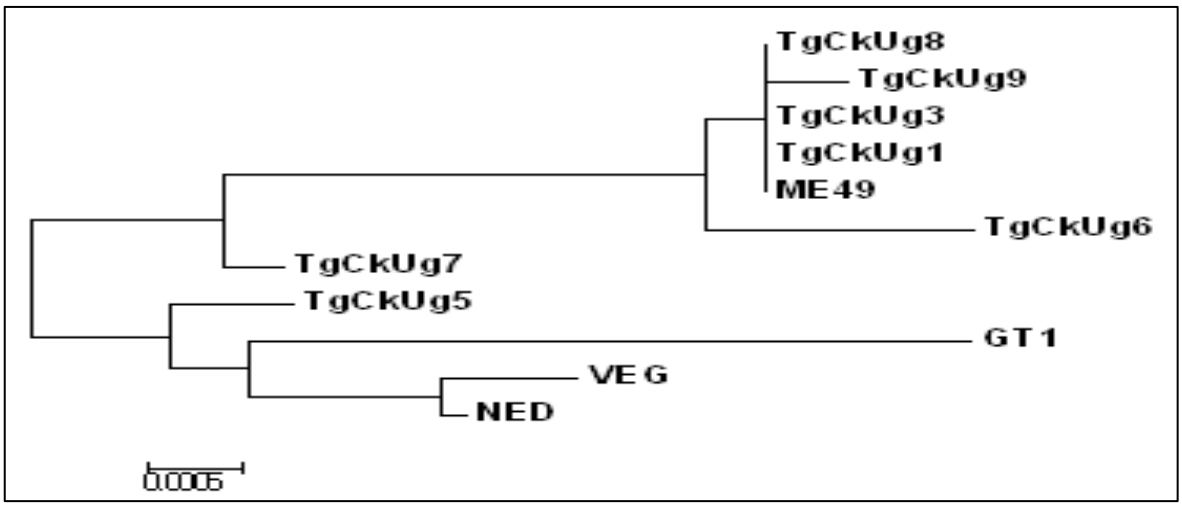

Figure 2.8: Construction of Neighbor-Joining tree using concatenated DNA sequences of SAG1, SAG2, BTUB, GRA6, L358, C22-8, C29-2, PK1 and Apico loci from NED, Ugandan strains (TgCkUg1, 3, 5, 6, 7, 8 and 9) and the reference strains of clonal type I, II and III (GT1, ME49, and VEG respectively)

TgCkUg1, 3 and 8 strains were shown to be identical to the reference ME49 strain, while the other Ugandan strains were differentiated from each other. It is obvious that these nine genetic markers have the ability to detect some variations between these local isolates but were still not able to totally differentiate them from each other and from the reference strain. However, by comparison, from this multi-locus PCR sequencing method using different nine markers and the previous multi-locus PCR-RFLP using 5 markers to analyze the same set of isolates (Bontell et al., 2009) (Table 2.9 and Figure 2.8), it is clear the improvement of discrimination power to identify the genetic variations between these local strains.

By examining phenotype variations between the Ugandan strains in Table 2.5, it is clear that three of these strains (TgCkUg1, 3 and 8) show high growth rate *in vitro* compared to the other Ugandan strains. Phylogenetic analysis of the sympatric Ugandan strains demonstrates that TgCkUg1, 3 and 8 strains are genotypically identical by using PCR sequencing technique, while the non-identical sympatric strains had lower growth rates, providing evidence that genotype may influence phenotype. Although TgCkUg1, 3 and 8 strains are

genotypically identical to the type II reference strain ME49, they demonstrate slower growth compared to ME49 strain, which reflect recent isolation of these sympatric Ugandan strains.

To evaluate the discrimination power of these genetic markers to differentiate more related type II strains, the DNA sequences of two type II strains (PRU and B73) and three type 12 strains (B41, ARI and RAY) of the 9 loci, retrieved from (http://www.toxodb.org), were added to the 7 Ugandan strains (Figure 2.9). Although most of these type II strains are differentiated by using the nine genetic markers, 4 of these strains are not distinguished from each other and are identical to the reference type II strain ME49 (TgCkUg1, 3, 8 and PRU).

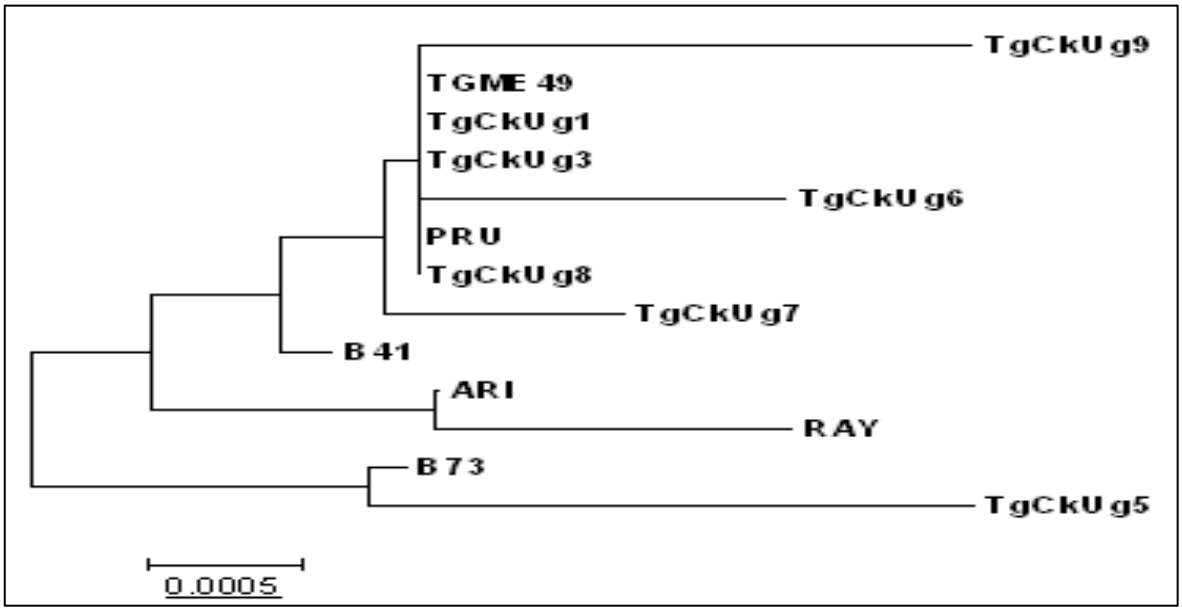

Figure 2.9: Construction of Neighbor-Joining tree using concatenated DNA sequences of SAG1, SAG2, BTUB, GRA6, L358, C22-8, C29-2, PK1 and Apico loci from the Ugandan strains (TgCkUg1, 3, 5, 6, 7, 8 and 9), the reference strain of clonal type II (ME49), type II strains (PRU and B73) and type 12 strains (B41, ARI and RAY) of *Toxoplasma gondii*

2.4 DISCUSSION

The parasite *Toxoplasma gondii* is an important pathogen with significant medical and veterinary importance; which is distributed worldwide and can infect all warm-blooded animals including humans. Genetically, it is quite conserved and thus r of intensive study is required to evaluate diversity in order to understand the population structure of this protozoan.

Previously, it was believed that *T. gondii* populations were precisely clonal. However, high genetic diversity confirmed through analysis of isolates from South America resulted in an unclear phylogenetic relationship between PCR-RFLP data from these isolates compared to those from Europe and North America. It is known that the results and characterizations of isolates achieved by applying PCR-RFLP technique is consistent when performed in areas with low genetic diversity, such as Europe and North America. However, when the genetic diversity is high in regions such as South America, the ability of the PCR-RFLP method to detect and describe the genetic variation of different isolates would be insufficient (Pena et al., 2008).

Multi-locus sequencing is believed to be one of the best methods used to analyze the parasite population structure and to identify polymorphism, as it is based on a large number of loci. Although PCR-RFLP methods are largely advanced as a result of revealing polymorphic restriction sites for endonucleases and they are cheaper and less time-consuming compared to sequencing, it is less informative than sequencing. This results from the assumption that all isolates will undergo the same pattern of restriction sites that were described previously for a few strains.

While the multi-locus PCR-RFLP technique is cheap and easily applied, it was originally established to determine the polymorphic sites that differentiate archetypal strains which circulate in North America and Europe. Therefore, it fails to determine unique polymorphisms that exist in the majority of alleles in strains isolated from regions with high genetic diversity such as South America (Pena et al., 2008). Otherwise, the reported multi-locus genotypes would be mistaken and misinterpreted as either archetypal or recombinant between archetypal strains.

In one study performed by Frazao-Teixeira et al. (2011) PCR-RFLP analysis at 10 genetic loci of five *T. gondii* isolates from pigs in Brazil revealed four genotypes. One of the

genotypes was archetypal and two isolates were similar. By applying DNA sequencing on the same isolates across the 10 PCR-RFLP loci examined, the relationship was resolved and 15 atypical alleles were identified, compared to the two alleles detected by multi-locus PCR-RFLP analysis.

A second study by Bezerra et al. (2012) applied multi-locus PCR-RFLP with seven genetic markers (SAG1, SAG2, SAG3, BTUB, C22-8, PK1 and Apico) to 11 isolates from pigs in Brazil and detected six different genotypes. Six of the isolates were indistinguishable by this method, creating a single genotype, while the other five isolates were characterized as individual genotypes. However, when multi-locus PCR sequencing was used through five different molecular markers (SAG1, SAG2, SAG3, BTUB and C22-8), all 11 strains were revealed to be different.

The application of DNA sequencing, then, is essential to improve the genetic characterization of atypical isolates. By comparison of PCR-RFLP and DNA sequencing results, it was concluded that the application of multi-locus PCR-RFLP may underestimate the true genetic diversity of *T. gondii* population (Frazao-Teixeira et al., 2011; Khan et al., 2006b). Therefore, multi-locus DNA sequencing is considered to be the method of choice to understand the real genetic diversity of *T. gondii* strains from South America.

In this study, however, in the level of variation across multiple loci relative to reference type I, type II and III strains, a high level of sequence homology was seen between the African isolates and the reference strains that originated from North America. Although genetic variations between these isolates were limited, these variations are restricted in relation to type II and III reference strains. From this multi-locus PCR sequencing method using different nine markers and the previous multi-locus PCR-RFLP using 5 markers to analyze the same set of isolate (Bontell et al., 2009) (Table 2.9 and Figure 2.8), the improvement of discrimination power to identify the genetic variation between these local strains is evident. Only 5 polymorphisms were identified in the multi-locus PCR-RFLP which represented the enzyme restriction sites, while 34 SNPs were detected by our multi-locus PCR sequencing method which improved the discrimination power of these markers to detect the variation among the local Ugandan strains.

Bontell et al. (2009) showed that TgCkUg5 strain was type II and TgCkUg6 was type III strain (Table 2.5). In our study, TgCkUg5 has mostly type III chromosomes, while TgCkUg6

mostly type II (Table 2.11 and Figure 2.6). Table 2.12 shows a comparison between the current study and previous two studies using PCR RFLP and PCR sequencing techniques for TgCkUg5 and 6 strains. In general, PCR sequencing of SAG1 and GRA6 for TgCkUg5 and 6 in the current were in agreement with Bontell, et al. (2009) study. In SAG2 locus, it is clear that both TgCkUg5 and 6 strains from chickens have mixture of types I and II by PCR RFLP analysis (Lindstrom, et al., 2007), while only one strain was retrieved in culture by PCR sequencing as type II in TgCkUg6 and type III in TgCkUg5 (while SAG2 cannot distinguish between type I and II). The additional markers in the current study (L358, PK1, C22-8 and C29-2) were type III in TgCkUg5 and type II in TgCkUg6.

Table 2.12: A comparison between the current study and previous two studies using PCR RFLP and PCR sequencing techniques for TgCkUg5 and 6 strains

	Previous studies							**Current study**							
	PCR RFLP from **chickens** (Lindstrom, et al., 2007)						PCR SEQ from **culture** (Bontell, et al., 2009)	PCR SEQ from **culture**							
Loci	SAG1	SAG2	SAG3	GRA6	BTUB	Genotype	Genotype (SAG1, SAG3, GRA6)	SAG1	SAG2	GRA6	BTUB	L358	PK1	C22-8	C29-2
TgCkUg5	1+2/3	1 + 2	1 + 2	1 + 2	2	Mix I + II	II	II/III	III	II	u-1	III	III	III	III
TgCkUg6	1+u-1	1 + 2	1 + 2 + 3	1 + 2	3	Mix I + II + III	III	II/III	II	III	II	II	II	II	II

In Table 2.12, in SAG1 marker, it is obvious that TgCkUg6 has a mixture of types I and u-1 (unique allele) by PCR RFLP analysis (Lindstrom, et al., 2007), while in PCR sequencing analysis in Bontell, et al. (2009) and the current studies, it has type II/III as SAG1 cannot distinguish between type II and III. In addition, in GRA6 marker, PCR RFLP analysis for TgCkUg6 resulted in a mixture of types I and II (Lindstrom, et al., 2007), while PCR

sequencing resulted in type III genotype in Bontell, et al. (2009) and the current studies. This is an evidence that PCR sequencing method improved the discrimination power to detect the variations within this strain as an example.

In addition, a comparison has been held between the available sequence data in the GenBnak of 34 loci for TgCkUg5 & 6 strains from Bontell, et al. (2009) study with the reference type II (ME49) and III (VEG) strains to determine the identity percentage of each strain with the reference (Table 2.13). Although the majority of TgCkUg5 loci are mostly identical to type II reference (ME49) and TgCkUg6 loci are mostly type III, there are some exceptions. In the VI-13 locus, both TgCkUg5 and 6 are more similar to type III reference (VEG), and in the X-26 locus, both of Ugandan strains are identical to type II reference (ME49) and not type III (VEG).

Many previous studies have concluded that isolates from South America have recombinant genotypes that assume new combinations of the main clonal archetypal alleles across the 10 PCR-RFLP loci examined (Dubey et al., 2008; Pena et al., 2008). However, the high number of unique alleles that were detected by multi-locus DNA sequencing within 5 pig isolates genotyped in (Frazao-Teixeira et al., 2011) compared to multi-locus PCR-RFLP analysis of the same isolates revealed that the South American isolates would not be classified as (recombinants) of the archetypal strains, but rather as (atypical) as they owned new combinations of both unique and archetypal alleles.

The recent advent of multi-locus genotyping has greater resolution than PCR-RFLP to identify genetic variation between isolates. The strength of the technique is in allowing comparisons to be made between large numbers of isolates, so that the basic population structure of the parasite is revealed. In the most recent studies, a panel of 11 loci is used to discriminate strains and these reveal an association with 6 major clades and 15 major haplotypes. This underlying association between global strain types is comprehensive and likely to be robust, providing a reference point for new isolates.

One criticism of the MLST approach is that the selection of loci can bias phylogenetic reconstruction. Using current methods, a range of loci has been selected on the grounds of distribution across chromosomes (Khan et al., 2005; Su et al., 2010, Su et al., 2012), but they are a heterogeneous mix biased towards exons from genes encoding functionally important proteins. It would be possible to test the robustness of the relationships by adding data from

additional loci to see whether the relationships hold.

Table 2.13: A comparison between the available sequence data in the GenBnak of 34 loci for TgCkUg5 and 6 strains from Bontell, et al. (2009) study with the reference type II (ME49) and III (VEG) strains

34 loci	**Identities**			
Loci	TgCkUg5 – (ME49) II	TgCkUg6 – (ME49) II	TgCkUg5 – (VEG) III	TgCkUg6 – (VEG) III
20.m03896 (ROP18)	100%	94%	94%	100%
33.m02185 (toxofilin)	99%	84%	84%	100%
65.m01164	100%	98%	98%	100%
80.m05038	100%	95%	95%	100%
641.m01562	100%	91%	91%	99%
Ib-3	100%	99%	99%	100%
II-4	99%	99%	98%	99%
II-5F	99%	85%	85%	100%
II-5R	99%	79%	79%	99%
III-7	100%	100%	100%	100%
IV-8	100%	99%	99%	100%
IV-9	100%	99%	99%	100%
V-10	100%	99%	99%	100%
V-12	99%	99%	99%	99%
VI-13	91%	91%	95%	96%
VI-14	100%	100%	100%	100%
VIIa-16	98%	99%	98%	99%
VIIb-17	100%	100%	100%	100%
VIIb-18	100%	98%	98%	100%
VIII-19F	100%	99%	99%	100%
VIII-19R	100%	99%	99%	99%
VIII-20F	100%	100%	100%	100%
VIII-20R	100%	99%	99%	100%
VIII-21	100%	99%	99%	100%
IX-22	100%	100%	100%	100%
IX-24	100%	99%	100%	99%
X-25	100%	99%	99%	100%
X-26	100%	100%	95%	94%
XI-30	100%	100%	100%	100%
XII-31	99%	98%	98%	99%
XII-32	100%	100%	100%	100%
B17 (MS)	100%	100%	100%	100%
W35 (MS)	100%	98%	98%	100%
UPRT-1 intron	100%	100%	100%	100%

Although a low level of variation among Toxoplasma isolates is seen in Europe, North America, Africa and Asia, it cannot be assumed that the parasite genotype is unimportant. It

is simply necessary to look deeper into the genome to make valid comparisons. This can be accomplished by additional sequencing. In this study, it is planned to use whole genome sequencing of additional Ugandan isolates in comparative studies of Toxoplasma genomics.

A key aim is to understand whether local allelic variation occurs within type II isolates. The initial study by Bontell et al. (2009) identified unique SNPs in the II/III recombinant isolate TgCkUg2. By sequencing additional strains it will be possible to test whether these alleles are present in all local strains (high clonality) or whether additional variants are found. Comparisons can then be made across geographically distinct type II isolates (Europe/North America) to compare allelic variation across the range of type II isolates (sequence of additional isolates) (Su et al., 2012).

Not only is there a recombinant strain in Uganda, TgCkUg2, (Bontell et al., 2009), but there is also evidence of recombination among further isolates in the region. In this study, it was found that TgCkUg5 has type II background SNPs in GRA6 (Chr.X) loci, while SNPs within C22-8 (Chr. Ib), C29-2 (Chr.III), L358 (Chr. V), PK1 (Chr. VI) and SAG2 (Chr. VIII) loci has type III background. This pattern of SNPs distribution in these chromosomes is totally in agreement with the same chromosomes of the recombinant strain, TgCkUg2, performed by Bontell et al. (2009) (Table 2.11 and Figure 2.6). In addition, type II background SNPs are located in L358 (Chr. V), PK1 (Chr. VI), BTUB (Chr. IX) and GRA6 (Chr.X) loci of the TgCkUg7 strain, while C22-8 (Chr. Ib), C29-2 (Chr.III) and SAG2 (Chr. VIII) loci has type III background SNPs. Again, this SNPs pattern is similar to that in TgCkUg2 except for one chromosome, V, which has type III background in TgCkUg2 strains (Table 2.11 and Figure 2.6). This pattern of SNPs distribution in these chromosomes is evidence of recombination between type II and III within these strains (TgCkUg5, 6 and 7).

It is clear that strains vary in their disease phenotype. For example, the growth rate of Ugandan strains *in vitro* shows considerable variation (Bontell et al., 2009) (Table 2.5). The high growth rate phenotype in TgCkUg1, 3 and 8 strains are correlated with genotype of these isolates, as they are identical, while the non identical sympatric strains had lower growth rates, providing evidence that genotype may influence phenotype. Although TgCkUg1, 3 and 8 strains are genotypically identical to the type II reference strain ME49, they demonstrate slower growth compared to ME49 strain, which reflect recent isolation of these sympatric Ugandan strains.

The risk of the sequence approach is that isolates will turn out to be truly clonal (no genetic variation). However even this outcome would be of value, as it would provide some of the first epidemiological evidence of clonality, as at the moment comparisons between isolates are mainly made on a global scale. Also, if genetically identical isolates prove to have different growth phenotypes, this points to the importance of epigenetic regulation of gene expression in controlling the disease phenotype.

3. CHAPTER 3: Genome level analysis of Ugandan isolates TgCkUg8 and 9

3.1 INTRODUCTION

Genome sequencing has dramatically advanced the scope of biological research to understand the molecular biology of whole systems instead of focusing on only certain genes or proteins. In 1995, sequencing of the first bacterial genome was achieved and, due to the development in sequencing technologies, the human genome was completed only six years later (Forrester and Hall, 2014). During this time, many genome projects for human parasites were established. The genome of *Plasmodium falciparum*, published in 2002, was the first project initiated in 1996 (Gardner et al., 2002). This genome project was established against a background of rising drug resistance in malaria and it was considered an approach which could enhance research for the discovery of new drugs and vaccines. However, the first published genome of an eukaryotic parasite was in 2001 for the microsporidian parasite *Encephalitozoon cuniculi* (Heinz et al., 2012). Many parasite genome studies had been completed and published by 2005 and became important resources to improve research in the investigation of parasites and to better understand the parasite host relationships.

The first published genome study based on the technique of high-throughput sequencing was for *Mycoplasma genitalium* in 2005. This publication indicated the initiation of a new era based on techniques that would later be called (next generation sequencing) (Margulies et al., 2005). In comparison with Sanger sequencing which is considered as (first generation) technology, next generation sequencing (NGS) methods are cheaper and produce large quantities of data exceeding one billion short reads per run in some cases (Metzker, 2010). Examples of NGS technologies are Roche/454 and Illumina/Solexa techniques. The early drawback of NGS was the generation of very short reads of data, as little as 21 bases, compared to up to 1000 bases reads produced by the Sanger sequencing method. However, this disadvantage was overcome via rapid improvement in (NGS) technologies.

3.1.1. Genome structure and genome sequencing

It is essential to study the genome structure of *T. gondii* and how it is organized in order to understand how the cellular processes are regulated. Genomic DNA is bound with a complex of nuclear proteins to form chromatin. There are two forms, a gene-rich

decondensed chromatin (euchromatin) where transcription is enhanced and a gene-poor condensed chromatin (heterochromatin) which is refractory to transcription. Generally, euchromatin is the active chromatin, while heterochromatin represents the constitutive chromatin. In fact, these epigenetic related structures are preserved from yeast to human including, of course, *T. gondii* (Gissot et al., 2007; Gissot et al., 2012). In Toxoplasma, histone and histone variants are considered to be epigenetic markers of both chromatins (euchromatin and heterochromatin) (Dalmasso et al., 2009).

The heterochromatin region has two identified domains, the telomeres and the centromere. The telomeres are located at the ends of the chromosome and consisting of the telomeric repeats and the subtelomeric region (Telomeric Associated Sequence - TAS). The subtelomere consists of repetitive elements and, in some cases, subtelomeric genes (Ottaviani et al., 2008; Pryde and Louis, 1999).

The subtelomeric genes play a significant role in many protozoan pathogens, such as *Plasmodium falciparum, Trypanosoma brucei, Trypanosoma cruzi* and *Leishmania major*. The genomes of *Trypanosoma brucei* (Berriman et al., 2005), *Trypanosoma cruzi* (El-Sayed et al., 2005) and *Leishmania major* (Ivens et al., 2005) were sequenced and published in 2005. Although there are considerable differences in lifestyles of these parasites, they were shown to have a conserved core gene set and a high level of diversity in the subtelomeres, which included several of the surface antigen genes. This difference in lifestyles among these parasites was also reflected by the fact that intracellular parasites, *Trypanosoma cruzi* and *Leishmania major*, exhibited a higher level of similarity compared to the extracellular parasite, *Trypanosoma brucei*. This pattern of telomeric diversity and central genetic conservation had been shown in many species such as the fungal parasite *Pneumocystis carinii* (Stringer and Cushion, 1998) and the apicomplexan parasite *Babesia bovis* (Brayton et al., 2007).

Although the significance of, and evidence for, the role of both telomeres and subtelomeric regions on the surrounding genes expression is established for many pathogens, the telomeres of Toxoplasma have not been studied. This is probably due to the fact that the genes at chromosome ends have not been analysed in this parasite and because the majority of the current genomes do not have complete assembled chromosomal ends.

In pathogen genomes, it is a common characteristic for the genes of surface antigens to be

frequently located in subtelomeric regions. This feature may play a role in enabling higher recombination among members of these gene families. The capability of recombination between central genes is restricted, while subtelomeric genes have greater ability to recombine, which results in generation of high diversity level in these antigen genes (Barry et al., 2005).

In *T. brucei*, sequencing of the genome showed that the percentage of encoded genes of variable surface glycoproteins (VSGs) in the genome was only 5% as complete genes, the majority being pseudogenes, which might produce new variants via recombination (Marcello and Barry, 2007). In addition, this pattern has been shown in other protozoans such as *Babesia spp* (Brayton et al., 2007).

In *P. falciparum*, it is known that the VAR family which includes about 60 genes has a significant role in the production of antigenic variation. The sequencing of the complete genome of this parasite contributed to understanding this family more clearly, by showing that they were not randomly allocated but more likely to be localized in telomeres. Additionally, it had been revealed the association of these genes in the regulation of antigenic variation (Gardner et al., 2002).

One of the common characteristics of parasitism is the capability to get nutrition from the host, as genes are lost that are not required. This pattern is shown as a reduction in the size of the genome, as some obligatory intracellular parasites have shorter genomes compared to their relatives free living parasites (Sakharkar et al., 2004). For instance, reduction in genome sizes had been shown in two of the Microsporidian obligatory intracellular parasites, *Trachipleistophora hominis* (Heinz et al., 2012) and *Encephalitozoon cuniculi* (Katinka et al., 2001). In contrast, *Trichomonas vaginalis* genome is an exception to this genome reduction pattern, and is considered as the largest protozoan sequenced genome with a size of 160Mb (Carlton et al., 2007).

Although understanding the genomic structure of many parasites has improved via the generation of many genome sequences, the generation of a high quality genome reference is still lacking. For instance, the genome of *Trypanosoma brucei* was sequenced and reported with 30 contigs (Berriman et al., 2005), while the genome of the relative species, *Trypanosoma congolense*, was sequenced with 3181 contigs (Jackson et al., 2012). This increasing pattern is due to a reduction in the cost of generating whole genome sequence

data while the closing gap cost of raw assemblies is high. Recently, new technology has been applied to generate complete closed genomes from shotgun sequencing of bacteria, which is called single molecule sequencing (Koren et al., 2013). As parasite genomes are more complicated compared to bacterial ones, the development of this technology would facilitate generation of complete parasitic genomes at a reduced cost.

Genomic analysis was usually performed on only a single representative strain from a particular species. After development of next generation sequencing methods, sequencing of many strains of a particular species improved our understanding of the population structure. For example, whole genome sequencing had been applied to detect recombination in *Trypanosoma brucei rhodesiense* (Goodhead et al., 2013) or in *T. gondii*, where ancestral recombination events were indicated through the pattern of polymorphism at the genome level (Boyle et al., 2006).

3.1.2. *Toxoplasma gondii* - Original sequence data and reference strains

The generation of whole genome sequences was initially completed for ME49 strain (type II), a strain originally isolated from sheep in the USA in 1958 (Table 3.1) by The Institute for Genomic Research (TIGR) to initiate the *Toxoplasma gondii* Genome Project in 1999-2002.

Table 3.1: The reference strains of *Toxoplasma gondii* (Khan et al., 2006a)

Strain	Type	Date of isolation	Source	Host	Reference
GT1	I	1980	USA	Goat	(Dubey, 1980)
RH	I	1939	USA	Human	(Sabin, 1941)
ME49	II	1958	USA	Sheep	(Kaufman et al., 1958)
VEG	III	1988	USA	Human	(Parmley et al., 1992)

Sequencing trials were carried out by whole genome shotgun sequencing via Sanger sequencing of ME49, which was considered to be the reference genome for *T. gondii*. It was shown that the coverage of ~10X through the use of conventional sequence reads provided adequate data required for *T. gondii* de novo assembly into 670 large scaffolds that ranged in size from <6 kb to > 7 Mbp, with a total estimated genome size of about 65 Mbp (Khan et al., 2005). Subsequently, by using a genetic linkage map, these scaffolds were re-assembled into the 14 chromosomes that ranged in size from 1.8 Mbp (chromosome Ia) to >7 Mbp (chromosome X) by using 250 genetic markers which were spaced at about 300 kb regular intervals across the genome (Table 3.2).

Table 3.2: The sizes of chromosomes of ME49 genome (Khan et al., 2005)

Chromosomes	Size (bp)	Chromosomes	Size (bp)
Ia	1856182	**VIIa**	4 502 211
Ib	1 956 324	**VIIb**	5 023 822
II	2 343 157	**VIII**	6 923 375
III	2 470 845	**IX**	6 384 456
IV	2 576 468	**X**	7 418 475
V	3 147 601	**XI**	6 570 290
VI	3 600 655	**XII**	6 871 637
TOTAL = 61 645 498 bp			

This resulted in a genome map of 61 Mbp in size which represented about 95% of the estimated 65 Mbp genome size (Khan et al., 2005). The difficulty in assembling some scaffolds into the genome resulted from either the low quality of the sequence or the existence of repeats. Later, the genomic data of type II ME49 strain was sequenced again by using next generation sequencing technology (paired-end 454), by J. Craig Venter Institute (JCVI), to generate a new higher quality genome and improve the annotation of this genome to be used as a reference for the other genomes. The coverage of this genome was improved to 26.55X (Tables 3.3 and 3.4).

Table 3.3: *Toxoplasma gondii* reference genomes (http://www.toxodb.org)

Chromosomes	Chromosomes length			
	GT1	**ME49**	**VEG**	**RH**
Ia	1841710	1859933	1874844	1986653
Ib	1949725	1955354	1974527	2013089
II	2257027	2347032	2341721	
III	2361114	2532871	2499703	
IV	2527423	2686605	2697628	
V	3030196	3331915	3263790	
VI	3539197	3656745	3680711	
VIIa	4380543	4541629	4652733	
VIIb	4975702	5069724	5093952	
VIII	6899611	6970285	6955806	
IX	6105434	6327655	6372921	
X	7353086	7486190	7456791	
XI	6541237	6623461	6626631	
XII	6851637	7094428	7144635	
TOTAL	**60613642**	**62483827**	**62636393**	

This effort was accompanied by further sequence generation for the two smallest chromosomes – (chromosomes Ia and Ib) from the type I RH strain which derived from humans in the USA in 1939, which were completed at the Wellcome Trust Sanger Centre via the Sanger sequencing method (Khan et al., 2006a) (Table 3.1). Subsequently, the genomes

were generated for the other reference strains type I (GT1), an isolate derived from goats in the USA in 1980, and type III (VEG) from humans in the USA in 1988 (Table 3.1), through the whole genome shotgun sequencing by using Sanger technology, in addition to the use of the paired-end illumina sequencing method, assembled and annotated by the J. Craig Venter Institute (JCVI), (Table 3.4).

Table 3.4: The whole genomes of the three reference strains of *Toxoplasma gondii* (http://www.toxodb.org & http://www.jcvi.org)

Reference strain	**Type**	**Sequencing technology**	**Chromosomes**	**Coverage**	**Genome & annotation source**
ME49	II	Sanger technology	14	10X	The Institute for Genomic Research (TIGR)
		Paired-end 454 GS FLX Titanium		26.55X	J. Craig Venter Institute (JCVI)
RH	I	Sanger technology	2 (Ia & Ib)		Wellcome Trust Sanger Institute
GT1	I	Sanger sequencing and Paired-end Illumina sequencing technologies	14	67.44x	J. Craig Venter Institute (JCVI)
VEG	III		14	77.38X	

The assembled and annotated whole genome sequences of these reference strains of *T. gondii,* in addition to community organizers and sequencing centres, which contributed in these projects, are available in a genome database for *T. gondii* http://www.toxodb.org/toxo/ (Gajria et al., 2008). This database ToxoDB is part of the Eukaryotic Pathogen Genome Database (EuPathDB).

3.1.3. Generation of additional genome sequence data for *Toxoplasma gondii* strains

Following the sequencing of the reference strains, and with the emerging knowledge of strain variation from the use of MLST and microsatellites, researchers have expanded the range of isolates sequenced through the generation of additional genome data. In 2009, Bontell et al., (2009) had sequenced the whole genome of the first natural recombinant strain from Africa, TgCkUg2, by using the 454 platform sequencing method, which was considered as one of the first whole genome sequencing projects to use next generation sequencing technology for SNP discovery. This sequencing method resulted in a genome sequence length of 51.8 Mbp which spanned over 84% of the genomic sequence of ME49 *T. gondii* genome. Despite low 4x coverage of 61,6 Mbp ME49 *T. gondii* genome, this was still

sufficient to identify SNPs by alignment and comparison with the previous sequenced reference strains of *T.gondii*.

In order to understand the population structure of *T. gondii,* it is necessary to generate new whole genome sequences from the representative members of the major lineages. In 2010, a study called the *Toxoplasma gondii* GSCID Project, had been established to generate high (20x) coverage of whole genome sequences and annotation for 13 new prototypic strains to be added to the previously sequenced whole genomes of the 3 reference strains (GT1, ME49 and VEG), to provide annotated genomes for reference strains for all 15 major lineages of *T. gondii* (Table 3.5) (https://sites.google.com/site/toxoplasmagondiigscidproject/home).

Table 3.5: The whole genomes of 13 new prototypic strains of *Toxoplasma gondii* using paired-end 454 GS FLX Titanium and paired-end Illumina sequencing technologies (http://www.toxodb.org & http://www.jcvi.org)

No.	Strain	Source	Host	Genome version	Length (Mbp)	Coverage
1.	ARI	USA	Human	01/03/2012	63.08	46.32X
2.	CAST	USA	Human	27/03/2012	63.05	74.24X
3.	COUG	Canada	Cougar	07/02/2013	63.7	46.75X
4.	CtCo5	Columbia, Bogota	Cat	17/07/2012	62.62	54.92X
5.	FOU	France	Human	23/06/2011	61.9	31.05x
6.	GAB2-2007-GAL-DOM2	Gabon, Makokou	Chicken	08/08/2012	62.98	98.92X
7.	MAS	France	Human	23/06/2011	61.48	42.75X
8.	RUB	French Guiana	Human	31/08/2011	62.61	30.56x
9.	TgCATBr5	Brazil	Cat	26/08/2011	61.64	31.55X
10.	TgCATBr9	Brazil	Cat	23/06/2011	61.82	24.37x
11.	TgCatPRC2	China	Cat	27/03/2012	62.98	28.52x
12.	VAND	French Guiana	Human	23/06/2011	62.33	Unknown
13.	p89	USA	Pig	23/06/2011	61.88	26.85x

Next generation sequencing techniques - (paired-end 454 GS FLX Titanium and paired-end Illumina sequencing) - were applied. Additionally, generation of moderate-coverage (10-12x coverage) whole genome sequences for 47 strains from all the 15 lineages were obtained by application of paired-end Illumina sequencing technology to evaluate and understand the population structure of *T. gondii* (Table 3.6). All sequencing, assembly and annotation were carried out in the J. Craig Venter Institute (JCVI). Next generation sequencing techniques (paired-end 454 GS FLX Titanium and paired-end Illumina sequencing) were used.

Table 3.6: 47 strains that are considered to be the representative members of the major lineages of *Toxoplasma gondii*, have been sequenced by paired-end illumina sequencing (http://www.toxodb.org)

NO.	STRAINS	Source	Host
1.	B41	USA	Bear
2.	B73	USA	Bear
3.	BOF	Belgium	Human
4.	BRC_TgH_18001	French Guiana	Human
5.	BRC_TgH_18002	French Guiana	Human
6.	BRC_TgH_18003	French Guiana	Human
7.	BRC_TgH_18009	French Guiana	Human
8.	BRC_TgH_18021	French Guiana	Human
9.	BRC_TgH_20005	Europe	Human
10.	BRC_TgH_21016	Europe	Human
11.	BRC_TgH_26044	Europe	Human
12.	CASTELLS	Uruguay	Sheep
13.	GAB1-2007-GAL-DOM10	Gabon, Dienga	Chicken
14.	GAB3-2007-GAL-DOM2	Gabon, Libreville	Chicken
15.	GAB3-2007-GAL-DOM9	Gabon, Libreville	Chicken
16.	GAB5-2007-GAL-DOM1	Gabon, Franceville	Chicken
17.	GAB5-2007-GAL-DOM6	Gabon, Franceville	Chicken
18.	GUY-2003-MEL	French Guyana	Human
19.	GUY-2004-ABE	French Guyana	Human
20.	GUY-2004-JAG1	French Guiana	Jaguar
21.	M7741	USA	Sheep
22.	PRU	France	Human
23.	RAY	USA	Human
24.	RH-88	USA, OH	Human
25.	RH-JSR	USA, OH	Human
26.	ROD	USA, CA	Human
27.	SOU	USA	Human
28.	TgCatBr1	Brazil	Cat
29.	TgCatBr10	Brazil	Cat
30.	TgCatBr15	Brazil	Cat
31.	TgCatBr18	Brazil	Cat
32.	TgCatBr25	Brazil	Cat
33.	TgCatBr26	Brazil	Cat
34.	TgCatBr3	Brazil	Cat
35.	TgCatBr34	Brazil	Cat
36.	TgCatBr44	Brazil	Cat
37.	TgCatBr64	Brazil, Sao Paulo	Cat
38.	TgCatBr72	Brazil, Sao Paulo	Cat
39.	TgCatPRC3	China	Cat
40.	TgCkBr141	Brazil, Para	Chicken
41.	TgCkCr1	Costa Rica	Chicken
42.	TgCkCr10	Costa Rica	Chicken
43.	TgCkGy2	Guyana	Chicken
44.	TgDogCo17	Columbia	Dog
45.	TgH26044	Europe	Human
46.	TgRsCr1	Costa Rica	Toucan
47.	TgShUS28	USA	Sheep

3.1.4. SNP mapping across genomes

Identification and mapping of SNPs across genomes is a powerful technique for evaluating genome structure and detecting recombination. Bontell et al. (2009) had identified the SNPs in the generated sequence of the African strain, TgCkUg2, through alignment with the published sequences for the reference strains of *T. gondii* ME49 (type II) and VEG (type III) and mapping them to the 14 chromosomes. The distribution of SNPs indicated that each of the chromosomes was inherited from a single parental strain, without evidence of chromosomal recombination –(chromosome sorting). Six chromosomes (II, IV, VI, VIIa, IX and X) originated from a type II parent strain (where TgCkUg2 is identical to ME49, but different from VEG), while 7 chromosomes (Ia, Ib, III, V, VIIb, VIII and XII) originated from type III strain (where TgCkUg2 is identical to VEG, but different from ME49). This was clear evidence that TgCkUg2 was a progeny of a cross between type II and III strains, without chromosomal recombination.

A study conducted by Boyle et al. (2006) mapped SNPs in different chromosomal regions via a comparison of the whole genome sequences of the type I, II and III references (GT1, ME49 and VEG respectively). It was shown that large chromosomal regions were dominated by only one of the three SNP types: (type I, II and III). For example, 90% of SNPs on chromosome IV were type III SNPs, while 98.7% of those on chromosome XI were of type I SNPs, and 67.2% of SNPs on chromosome X were type II. Chromosome Ia had obviously fewer SNPs than the other chromosomes. In addition, the pattern of SNPs in the genome of *T. gondii* had been used to construct a model of the genetic history of the clonal type I, II and III of *T. gondii*. It was clear that each chromosome consisted of a maximum of two different types of SNPs regions, which suggested the occurrence of few ancestral recombination crosses between distinct genotypes (Figure 3.1).

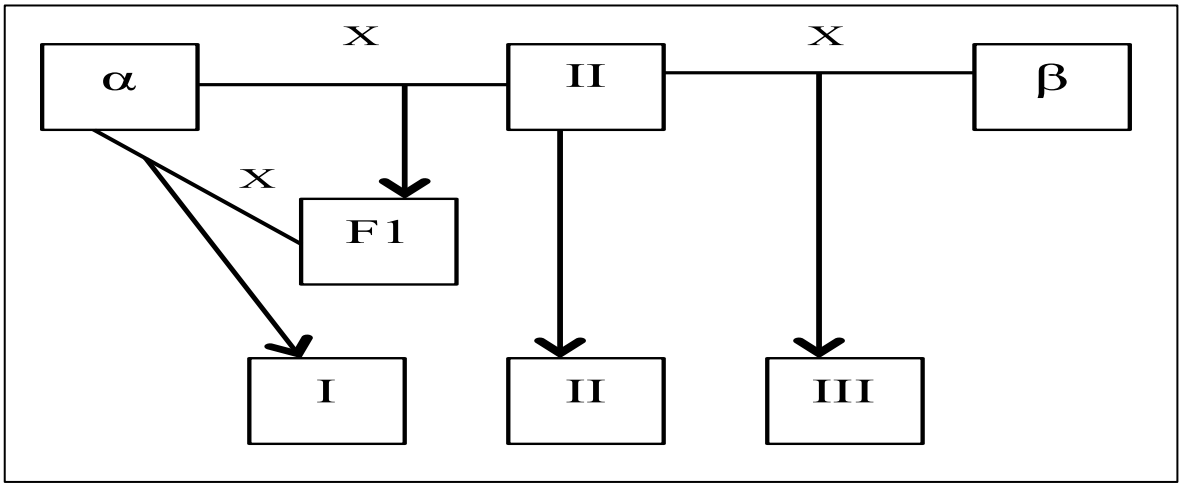

Figure 3.1: Model for the genetic history of the clonal type I, II and III strains of *Toxoplasma gondii*. It is suggested that type III strain resulted from a single cross between type II and a hypothetical strain named β. A single cross between type II and a hypothetical strain called α had given rise to F1 progeny which was followed by another cross between F1 progeny and α strain to give rise to type I strain (adopted from (Boyle et al., 2006))

This broader scale analysis of SNP's provides a useful method of detecting and mapping recombination (Boyle et al., 2006). Bontell et al. (2009) used this method to map type I SNP-dominated regions in TgCkUg2, and showed regions of high similarity between type II and III regions in Boyle's earlier analysis.

3.1.5. Aims

To generate the whole genome sequences of two of type II *T. gondii* strains with different phenotypes (growth rate in tissue culture), TgCkUg8 (high growth rate) & TgCkUg9 (low growth rate), using the Illumina MiSeq Paired-End Sequencing and to conduct a comparison with the type II reference strain (TgME49) through alignment to estimate the level of local variation within type II strains, in addition to look for further evidence of recombination in African strains other than TgCkUg2 (Bontell et al., 2009).

3.2 METHODS

3.2.1 Cell culture

Madin-Darby Bovine Kidney (MDBK) cell line was used through applying the same protocol that mentioned in Section 2.2.2.

3.2.2 Parasite culture

Tachyzoites of the TgCkUg8 and 9 strains of *T. gondii* were cultured by applying the same protocol that mentioned in Section 2.2.3. The released tachyzoites in the supernatant were transferred to a centrifuge tube and spun down for 10 min at 2500 r.p.m. This low speed ensured that dense parasites were collected and bovine cells excluded. The parasites were then washed three times by PBS (3000 rpm, 5 min) to remove cell debris. The remaining intracellular parasites were released by passage through 20g needles.

3.2.3 DNA extraction and purification

DNA from culture-derived tachyzoites of *T. gondii* isolates (TgCkUg8 and 9) was extracted and purified by using the Qiagen DNeasy Blood and Tissue Kit (see Section 2.2.4). The measurement of volumes, concentrations and confirmation of purity (by values of >= 1.8 for the Nanodrop 260/280 ratio) of the purified DNA for both samples was applied by using Nanodrop, prior to being sent for paired-end Illumina sequencing which was carried out by Source Bioscience (Table 3.7).

Table 3.7: The concentrations and confirmation of purity of the purified DNA for TgCkUg8 and 9 samples

Samples (DNA)	**Nucleic acid concentration (ng/ µl)**	**260/280**
TgCkUg8	58.6	1.94
TgCkUg9	73.4	1.98

3.2.4 Next generation sequencing (paired-end Illumina sequencing)

The purified DNA of samples TgCkUg8 and 9 were sent for paired-end Illumina sequencing which was carried out by Source Bioscience. The samples passed the initial QC (quality control), as the concentrations and volumes required were sufficient for the 550bp library prep.

3.2.4.1 Preparation of the DNA Libraries

The Illumina TruSeq DNA Nano library preparation was performed to prepare the samples for Illumina paired-end multiplexed sequencing. The libraries were loaded at a concentration of 8pM onto 1 lane of an Illumina MiSeq flow cell. Samples were then sequenced using 250bp paired-end run.

3.2.4.2 Raw reads data

The raw sequence data of samples TgCkUg8 and 9 were provided as text files in FASTQ format. For paired-end sequencing, two files were provided per sample, one file containing forward reads, and another file containing reverse reads.

3.2.4.3 Mapping of raw reads

Mapping of raw reads to the published type II reference strain genome of *Toxoplasma gondii* (TgME49) (http://toxodb.org/common/downloads/Current_Release/TgondiiME49/fasta/data/), was carried out by Source Bioscience using BWA (version: 0.7.5a-r405).

The raw reads for each sample (TgCkUg8 and 9) were mapped with the reference sequence and stored in SAM format files. SAM stands for (Sequence Alignment Map). The SAM files were then converted to BAM format files, the binary equivalent of SAM but in a compressed version. The BAM files were then sorted and indexed for each of the Ugandan samples TgCkUg8 and 9 (Figure 3.2).

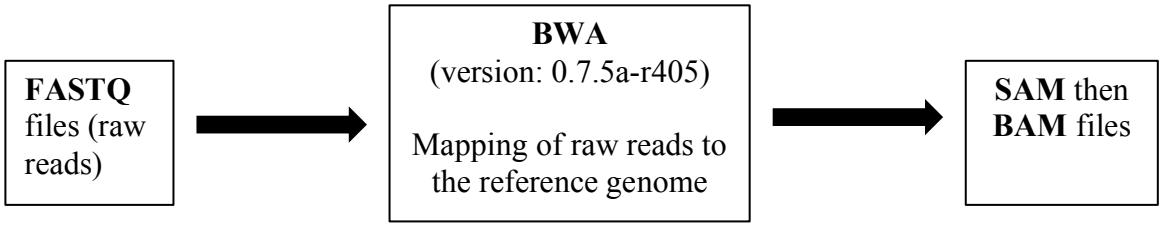

Figure 3.2: Summary of mapping of raw reads against the reference genome

3.2.4.4 SNP calling

The mutation analysis was performed by Dr. Rachel Brenchley at the University of Salford, using BCFtools (http://samtools.github.io/bcftools/). A customised set of commands was used to call SNPs in TgCkUg8 and 9 mapped reads against the Type II reference strain (TgME49) with 3 and 5 reads depths for each sample (see Appendix 1). The SNPs were

stored in VCF (Variant Call Format) files for both TgCkUg8 and 9 datasets. VCF is a standardised text file format for representing SNP, indel, and structural variation calls.

3.2.4.5 Visualization of the coverage and variations within TgCkUg8 and 9

The sorted and indexed BAM files and VCF files were loaded into IGV (Integrative Genomics Viewer) genome browsers (Version 2.3.49) (http://www.broadinstitute.org/igv/home) and Artemis (version16.0.0) (https://www.sanger.ac.uk/resources/software/artemis/) for visualization of the coverage and distribution of variant loci across the 14 chromosomes of the type II reference strain genome of *T. gondii* (TgME49).

3.3 RESULTS

3.3.1 Illumina Sequencing

3.3.1.1 Raw reads data

The raw sequence data of samples TgCkUg8 and 9 were provided as text files in FASTQ format which provides per-base quality scores to each read. These quality scores or (Q) reflect the confidence of accuracy of a given base-call (it reflects the probability that the called base is wrong). Briefly, the scores ranged from 2 to 40, where 40 is the highest effective score, indicating that there is a 1 in 10,000 chance that the called base is incorrect (99.99% accurate) (Table 3.8**)**. The Q30 threshold (indicating accuracy of 99.9%) is a standard indicator of read quality. Only reads with Q>=30 were aligned with the reference genome TgME49.

Table 3.8: Quality score and error probability of the raw data (FASTQ) (P=10^(-Q/10.0), P: probability, Q: quality.

Quality score (Q)	**Error probability (P) (Probability of incorrect base call)**	**Base call accuracy**
10	0.1 (1 in 10)	90%
20	0.01(1 in 100)	99%
30	0.001(1 in 1000)	99.9%
40	0.0001(1 in 10000)	99.99%

The total number of reads that were generated by paired-end sequencing with 250bp average length and which passed the quality control criteria is summarised in Table 3.9 below which shows the numbers of forward and reverse reads for both TgCkUg8 and 9 samples.

Table 3.9: A summary of the total number of reads generated by paired-end sequencing with 250bp average length for reads of TgCkUg8 &9

Samples	**Number of reads**		
	Total (paired) (QC-passed reads)	Read 1 (Forward)	Read 2 (Reverse)
TgCkUg8	13665575	6833149	6832426
TgCkUg9	12137544	6061063	6076481

3.3.1.2 Mapping of raw reads

Mapping of raw reads to the published type II reference strain genome of *Toxoplasma gondii* (TgME49) was carried out by Source Bioscience using BWA (version: 0.7.5a-r405). The

mapping quality was measured by a Phred score, which is identical to the quality measure in the FASTQ file (see Table 3.8). If Q=30, P=1/1000, on average, one out of 1000 alignments will be wrong.

The total number forward and reverse reads that had been mapped with the reference genome TgME49 for all chromosomes of TgCkUg8 &9 is summarized in Table 3.10. It is obvious that coverage is low and that only 5.53% of the total generated reads were mapped to the reference genome TgME49 for TgCkUg8, compared to 14.68% of the total generated reads that had been mapped against the reference genome for TgCkug9 sample.

Table 3.10: A summary of the number of mapped reads for all chromosomes of TgCkUg8 &9 with the reference genome TgME49

Chromosomes	Number of reads					
	TgCkUg8			TgCkUg9		
	Read 1 (Forward)	Read 2 (Reverse)	Total	Read 1 (Forward)	Read 2 (Reverse)	Total
Ia	7839	7516	15355	19867	19453	39320
Ib	9880	9516	19396	23072	22858	45930
II	12606	11896	24502	26442	25698	52140
III	13247	12883	26130	32885	32634	65519
IV	15542	14929	30471	35849	34515	70364
V	19732	19389	39121	50399	50054	100453
VI	20955	20478	41433	45669	45529	91198
VIIa	23879	22202	46081	62825	61013	123838
VIIb	24526	23977	48503	51684	50879	102563
VIII	38779	38017	76796	101254	100565	201819
IX	50407	46564	96971	111482	107339	218821
X	62656	62522	125178	138086	136969	275055
XI	37445	37492	74937	89018	89297	178315
XII	45336	45425	90761	108182	107986	216168
TOTAL	**382829**	**372806**	**755635 (5.53%)**	**896714**	**884789**	**1781503 (14.68%)**

The total coverage of these mapped reads of the reference genome TgME49 is estimated at read depths of 3 and 5 for all chromosomes of TgCkUg8 and 9 (Table 3.11).

Table 3.11: The coverage for all chromosomes for TgCkUg8 and 9 at 3x and 5x depths against the reference genome TgME49.

Chromosomes	Chr. Length ($x10^5$)	TgCkUg8				TgCkUg9			
		At 3x depth		At 5x depth		At 3x depth		At 5x depth	
		No. bases covered ($x10^5$)	Percent coverage %	No. bases covered ($x10^5$)	Percent coverage %	No. bases covered $x10^5$	Percent coverage %	No. bases covered ($x10^5$)	Percent coverage (%)
Ia	18.6	2.1	11.5	0.3	1.3	9.8	52.9	4.8	25.6
Ib	19.6	2.4	12.1	0.3	1.8	10.8	55.1	5.6	28.6
II	23.5	2.8	11.9	0.4	1.5	12.1	51.5	5.5	23.6
III	25.3	2.9	11.5	0.4	1.7	14.4	56.7	7.7	30.3
IV	26.9	3.3	12.4	0.5	1.9	14.4	53.5	7.2	26.6
V	33.3	4.2	12.6	0.7	2.1	18.6	55.7	9.6	28.8
VI	36.6	4.4	12.1	0.7	1.8	19.1	52.2	8.9	24.6
VIIa	45.4	5.3	11.7	0.7	1.6	27.5	60.5	15.9	35.2
VIIb	50.7	5.9	11.7	0.8	1.6	25.2	49.7	10.8	21.3
VIII	69.7	8.1	11.7	1.1	1.6	41.9	60.1	23.9	34.3
IX	63.3	7.8	12.3	1.3	1.9	37.8	59.7	21.1	33.3
X	74.9	8.7	11.6	1.2	1.7	44.6	59.6	25.3	33.8
XI	66.2	7.9	11.9	1.1	1.6	39.6	59.7	21.6	32.6
XII	70.9	8.9	12.6	1.4	2.1	40.6	57.3	22.2	31.3
Total (Average)	**624.9**	**74.7**	**12.0**	**10.9**	**1.7**	**356.4**	**57.1**	**190.1**	**30.4**

The estimated total coverage of the reads generated by paired-end Illumina sequencing for TgCkUg8 is 12.0% and 1.7% of the reference genome TgME49 (62.5 Mbp) at 3x and 5x depths respectively. For TgCkUg9, the estimated total coverage is 57.15 and 30.4% of the reference genome TgME49, at 3x and 5x depths respectively. These coverage percentages are low in comparison with the total coverage of sequences generated by the 454 GSFLX method for TgCkUg2, which was 84.09% of the reference genome TgME49 (61.6 Mbp) at 4x depth (Bontell et al., 2009) (Figures 3.3 and 3.4). In both Ugandan strains, the coverage percentage for all chromosomes is mostly similar at 3 and 5 reads depths for each sample. This indicates that there is no obvious coverage bias for all chromosomes for both strains.

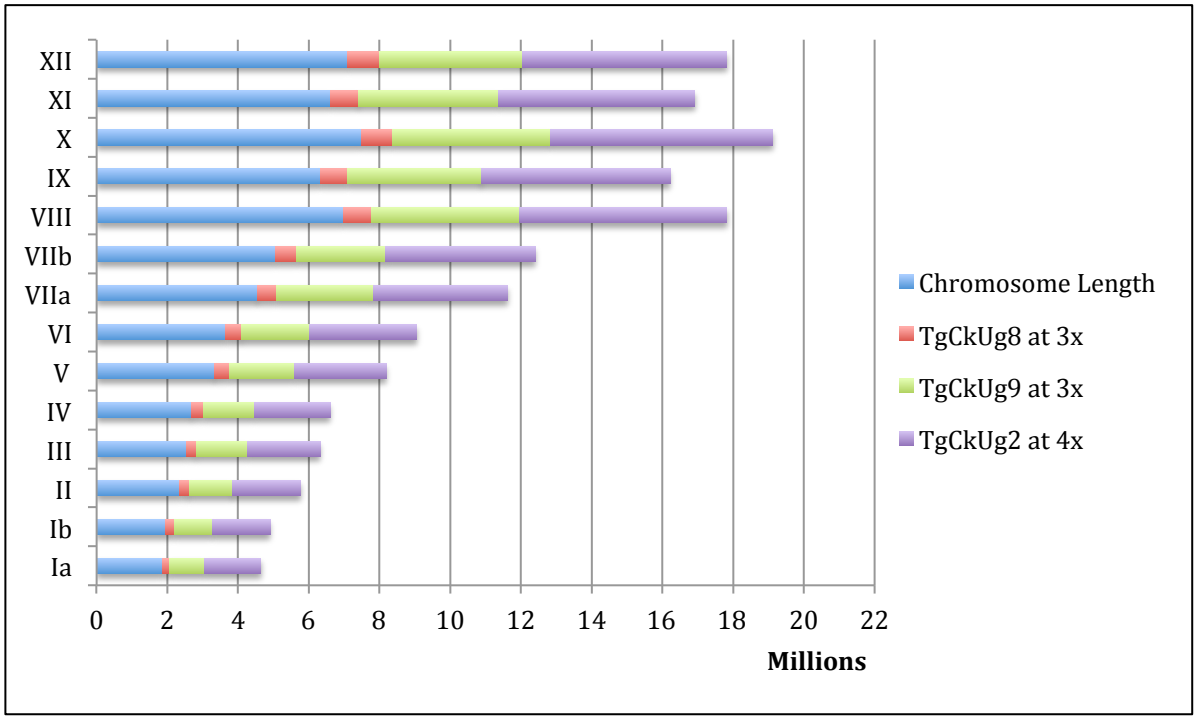

Figure 3.3: The coverage for all chromosomes of TgCkUg8 and 9 at 3x depth compared with the coverage of TgCkUg2 at 4x depth

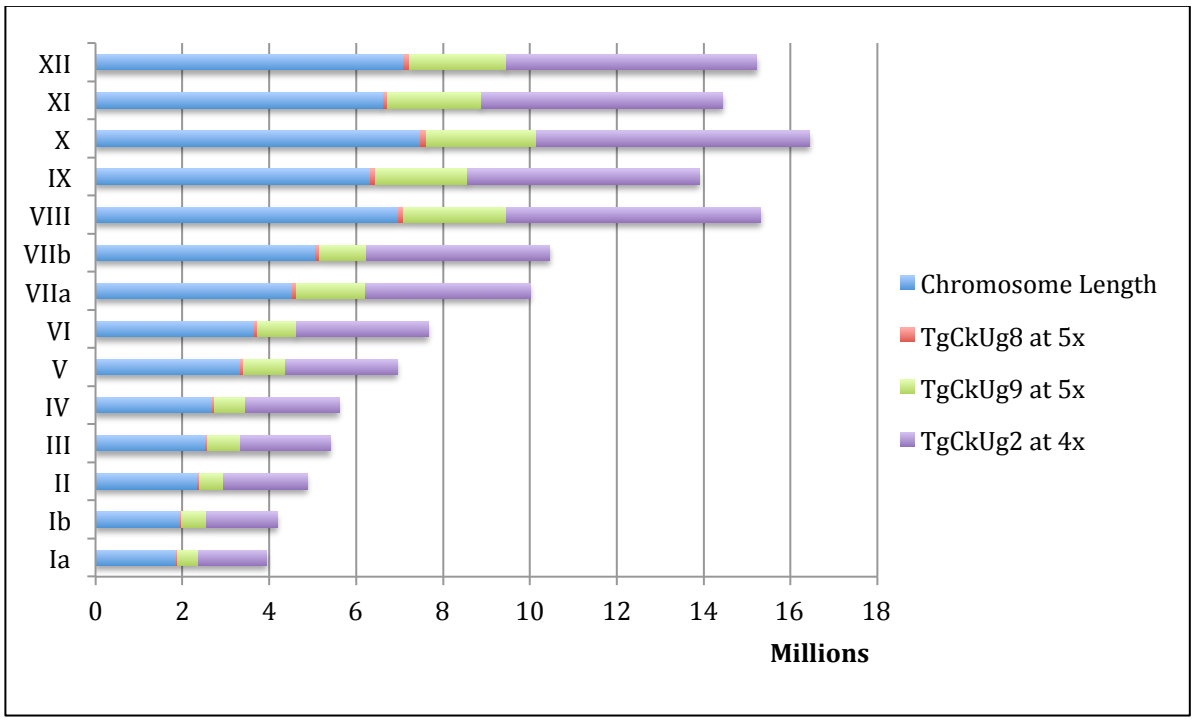

Figure 3.4: The coverage for the chromosomes of TgCkUg8 and 9 at 5x depth compared with the coverage of TgCkUg2 at 4x depth

In addition, the total coverage of genes, exons and introns was estimated with 3 and 5 read depths for all chromosomes of TgCkUg8 (Table 3.12) and 9 (Table 3.13).

Table 3.12: The coverage for the genes, exons and introns for all chromosomes of TgCkUg8 at 3x and 5x depths

Chromosomes	TgCkUg8								
	Total genic bp ($x10^5$)	Total exonic bp ($x10^5$)	Total intronic bp ($x10^5$)	No. bases covered at 3x ($x10^5$)			No. bases covered at 5x ($x10^5$)		
				Genes	**Exons**	**Introns**	**Genes**	**Exons**	**Introns**
Ia	15.5	9.8	5.7	1.8	1.1	0.7	0.2	0.1	0.1
Ib	16.3	9.2	7.1	1.9	1.1	0.8	0.3	0.1	0.2
II	18.4	10.2	8.2	2.2	1.1	1.1	0.3	0.1	0.2
III	18.6	11.6	7	2.1	1.2	0.9	0.3	0.2	0.1
IV	20.8	12.7	8.1	2.5	1.5	1	0.3	0.2	0.1
V	24.4	14.4	10	3.1	1.7	1.4	0.4	0.2	0.2
VI	29.6	18.6	11	3.6	2.1	1.5	0.5	0.3	0.2
VIIa	37.4	23.3	14.1	4.4	2.6	1.8	0.6	0.3	0.3
VIIb	35.2	18.5	16.7	4.2	2.1	2.1	0.6	0.2	0.4
VIII	48.7	26.9	21.8	5.8	3.1	2.7	0.8	0.3	0.5
IX	51.2	31.6	19.6	6.3	3.6	2.7	0.9	0.5	0.4
X	61.9	38.9	23	7.1	4.2	2.9	0.9	0.5	0.4
XI	54.9	33.6	21.3	6.5	3.8	2.7	0.8	0.4	0.4
XII	57.6	35.1	22.5	7.1	3.9	3.2	1.1	0.5	0.6
TOTAL	**490.5**	**294.4**	**196.1**	**58.6**	**33.1**	**25.5**	**8**	**3.9**	**4.1**

Table 3.13: The coverage for the genes, exons and introns for all chromosomes of TgCkUg9 at 3x and 5x depths

Chromosomes	TgCkUg9								
	Total genic bp ($x10^5$)	Total exonic bp ($x10^5$)	Total intronic bp ($x10^5$)	No. bases covered at 3x ($x10^5$)			No. bases covered at 5x ($x10^5$)		
				Genes	**Exons**	**Introns**	**Genes**	**Exons**	**Introns**
Ia	15.5	9.8	5.7	8.6	5.5	3.1	4.3	2.8	1.5
Ib	16.3	9.2	7.1	9.2	5.3	3.9	4.9	2.9	2
II	18.4	10.2	8.2	9.9	5.8	4.1	4.7	2.9	1.8
III	18.6	11.6	7	11.1	6.9	4.2	6.1	3.9	2.2
IV	20.8	12.7	8.1	11.6	7.2	4.4	5.9	3.6	2.3
V	24.4	14.4	10	14.3	8.7	5.6	7.7	4.7	3
VI	29.6	18.6	11	16.1	10.3	5.8	7.7	5.1	2.6
VIIa	37.4	23.3	14.1	23.3	14.8	8.5	13.8	8.8	5
VIIb	35.2	18.5	16.7	18.5	9.9	8.6	8.2	4.5	3.7
VIII	48.7	26.9	21.8	30.1	17.1	13	17.5	10.1	7.4
IX	51.2	31.6	19.6	31.4	19.6	11.8	17.6	11.1	6.5
X	61.9	38.9	23	38.1	24.4	13.7	21.8	14.3	7.5
XI	54.9	33.6	21.3	33.8	20.9	12.9	18.6	11.9	6.7
XII	57.6	35.1	22.5	34.1	20.9	13.2	18.9	11.7	7.2
TOTAL	**490.5**	**294.4**	**196.1**	**290.1**	**177.3**	**112.8**	**157.7**	**98.3**	**59.4**

3.3.1.3 SNP calling

The generated sequences (reads) for both TgCkUg8 and 9 were mapped to the type II reference genome of *Toxoplasma gondii* (TgME49) to identify SNPs in all 14 chromosomes. SNPs were identified at read depths of both 3 and 5 for all chromosomes of TgCkUg8 and 9 (Table 3.14).

Table 3.14: Number of SNPs in all chromosomes of TgCkUg8 & 9 at 3x and 5x depths and shared SNPs between TgCkUg8 & 9 compared with the reference genome TgME49

Chromosomes	**TgCkUg8**		**TgCkUg9**		**Shared SNPs (Between TgCkUg8 and 9)**	
	At 3x	**At 5x**	**At 3x**	**At 5x**	**At 3x**	**At 5x**
Ia	106	17	462	203	28	4
Ib	363	99	763	385	140	55
II	370	363	970	390	124	198
III	664	299	1653	936	322	171
IV	659	197	1138	588	201	54
V	1242	1077	2223	1287	766	535
VI	830	621	1170	565	553	240
VIIa	502	215	1305	823	250	152
VIIb	608	142	1454	568	208	69
VIII	1002	318	2372	1406	418	224
IX	1183	405	2576	1474	540	275
X	1668	1027	2624	1489	997	415
XI	1228	864	1733	937	727	277
XII	2149	949	4200	2473	1111	582
Total	**12574**	**6593**	**24643**	**13524**	**6385**	**3251**

The variation among both Ugandan strains compared to the reference strain is informed by identification of the total number of SNPs, called against the reference genome sequence. This was 12574 in TgCkUg8, and 24643 in TgCkUg9 at 3 reads depth. To increase the confidence in the SNPs, they were called against the reference genome sequence at 5 reads depth, with a total number of 6593 in TgCkUg8 and 13524 in TgCkUg9. The total number of SNPs that were shared between TgCkUg8 and 9 against the type II reference genome sequence of *Toxoplasma gondii* (TgME49) was 6385 at 3 reads depth, while it was 3251 at 5 reads depth, which is consistent with the level of variation between the two Ugandan strains being lower than with the type II reference strain TgME49.

Due to the relatively low coverage of the reference genome TgME49 for both Ugandan strain sequences, it is more appropriate to estimate SNP density (SNPs/kb) rather than total

SNPs number. Thus, the total estimated SNPs/kb in TgCkUg8 is 1.7 at 3 reads depth, while it is elevated to 6.1 at 5 reads depth, with clear variations of estimated SNPs/kb across 14 chromosomes at both reads depths (Table 3.15).

Table 3.15: Number of SNPs and estimated SNPs/kb in all chromosomes of TgCkUg8 at 3x and 5x depths compared with the reference genome TgME49.

Chromosomes	TgCkUg8					
	At 3x			At 5x		
	No. bases covered (x10^5)	No. of SNPs	**SNPs/kb**	No. bases covered (x10^5)	No. of SNPs	**SNPs/kb**
Ia	2.1	106	**0.5**	0.3	17	**0.6**
Ib	2.4	363	**1.5**	0.3	99	**3.3**
II	2.8	370	**1.3**	0.4	363	**9.1**
III	2.9	664	**2.3**	0.4	299	**7.5**
IV	3.3	659	**2.0**	0.5	197	**3.9**
V	4.2	1242	**2.9**	0.7	1077	**15.4**
VI	4.4	830	**1.9**	0.7	621	**8.9**
VIIa	5.3	502	**0.9**	0.7	215	**3.1**
VIIb	5.9	608	**1.1**	0.8	142	**1.8**
VIII	8.1	1002	**1.2**	1.1	318	**2.9**
IX	7.8	1183	**1.5**	1.3	405	**3.1**
X	8.7	1668	**1.9**	1.2	1027	**8.6**
XI	7.9	1228	**1.6**	1.1	864	**7.9**
XII	8.9	2149	**2.4**	1.4	949	**6.8**
Total	**74.7**	**12574**	**1.7**	**10.9**	**6593**	**6.1**

In TgCkUg9, the total estimated SNPs/kb is 0.7 at both 3 and 5 reads depths, which is the same value in TgCkUg2 at 4 reads depth (Table 3.16**)**. This variation in the estimated SNPs/kb in TgCkUg8 at both 3 and 5 reads depths may be explained by the very low coverage percentages, with 12.0% and 1.7% of the reference genome TgME49 (62.5Mbp) at 3x and 5x depths respectively. The percentage coverage is higher in TgCkUg9 with 57.15% and 30.4%, of the reference genome TgME49, at 3x and 5x depths respectively. The distribution of the number of SNPs and estimated SNPs/kb across the 14 chromosomes in both Ugandan strains with both 3 and 5 reads depths are summarized in Tables 3.14, 15 and 16.

Table 3.16: Number of SNPs and estimated SNPs/kb in all chromosomes of TgCkUg9 at 3x and 5x depths compared with the reference genome TgME49

Chromosomes	TgCkUg9					
	At 3x				At 5x	
	No. bases covered (x10^5)	No. of SNPs	**SNPs/kb**	No. bases covered (x10^5)	No. of SNPs	**SNPs/kb**
Ia	9.8	462	**0.5**	4.8	203	**0.4**
Ib	10.8	763	**0.7**	5.6	385	**0.7**
II	12.1	970	**0.8**	5.5	390	**0.7**
III	14.4	1653	**1.1**	7.7	936	**1.2**
IV	14.4	1138	**0.8**	7.2	588	**0.8**
V	18.6	2223	**1.2**	9.6	1287	**1.3**
VI	19.1	1170	**0.6**	8.9	565	**0.6**
VIIa	27.5	1305	**0.5**	15.9	823	**0.5**
VIIb	25.2	1454	**0.6**	10.8	568	**0.5**
VIII	41.9	2372	**0.6**	23.9	1406	**0.6**
IX	37.8	2576	**0.7**	21.1	1474	**0.7**
X	44.6	2624	**0.6**	25.3	1489	**0.6**
XI	39.6	1733	**0.4**	21.6	937	**0.4**
XII	40.6	4200	**1.1**	22.2	2473	**1.1**
Total	**356.4**	**24643**	**0.7**	**190.1**	**13524**	**0.7**

To show the coverage and variations of TgCkUg8 and 9 sequences across the 14 chromosomes of the type II reference genome of *T. gondii* (TgME49), IGV genome browsers and Artemis were used. Figures 3.5, 3.6, 3.7 and 3.8 show that the density of SNPs is higher towards the ends of most of the chromosomes (telomeres).

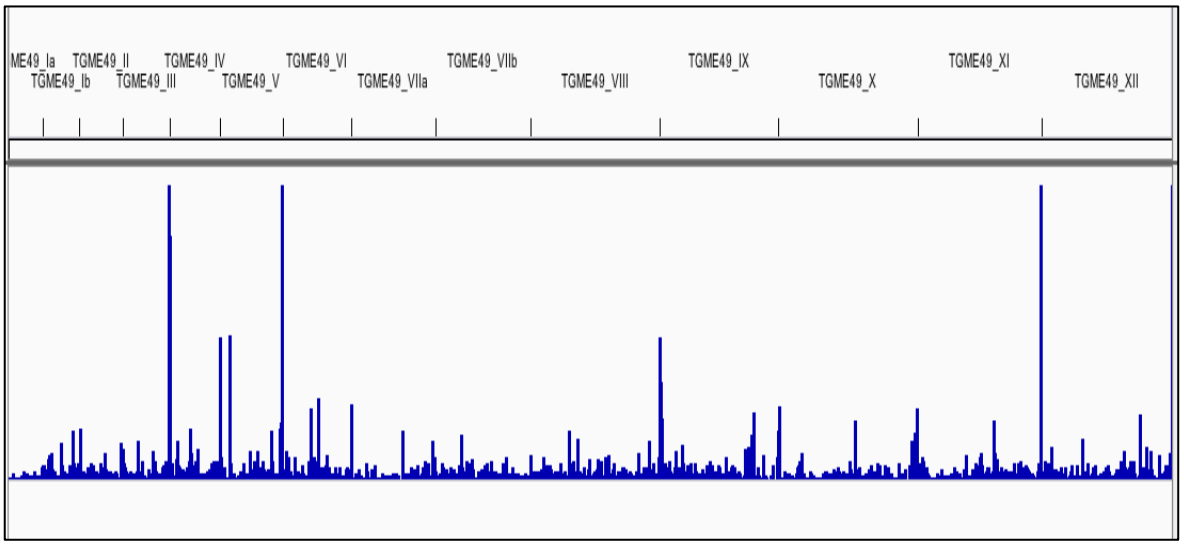

Figure 3.5: The variations of both Ugandan strains (TgCkUg8 & 9) compared to the reference genome across all 14 chromosomes of TgME49 (using IGV genome browsers)

By comparing graphs of the coverage and SNPs of both TgCkUg8 and 9 sequences in Figure 3.6, 3.7 and 3.8, it is possible to see the relationship between the coverage percentage of the Ugandan strain sequences and the number of SNPs in this strain. The higher total number of SNPs (24643 and 13524) is identified in TgCkUg9 which has the higher coverage percentage (57.15% and 30.4%) at both x3 and x5 depths respectively (Tables 3.11 and 3.16).

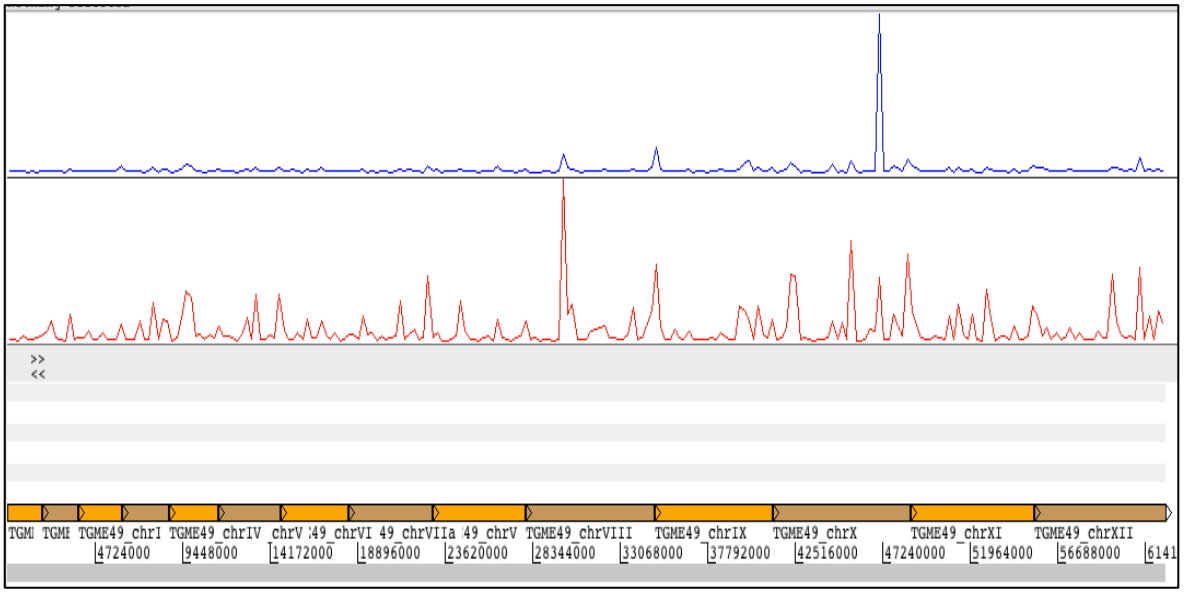

Figure 3.6: Coverage graph of TgCkUg8 (mapped reads) to the reference genome chromosomes (TgME49) in blue and SNPs graph in red (using Artemis)

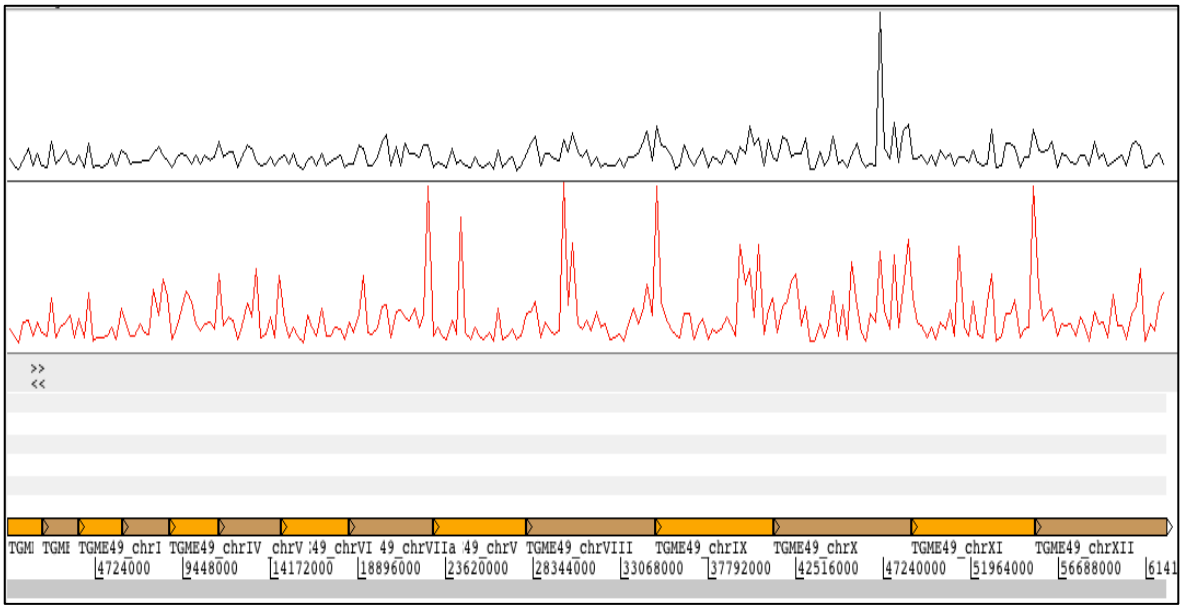

Figure 3.7: Coverage graph of TgCkUg9 (mapped reads) to the reference genome chromosomes (TgME49) in black and SNPs graph in red (using Artemis)

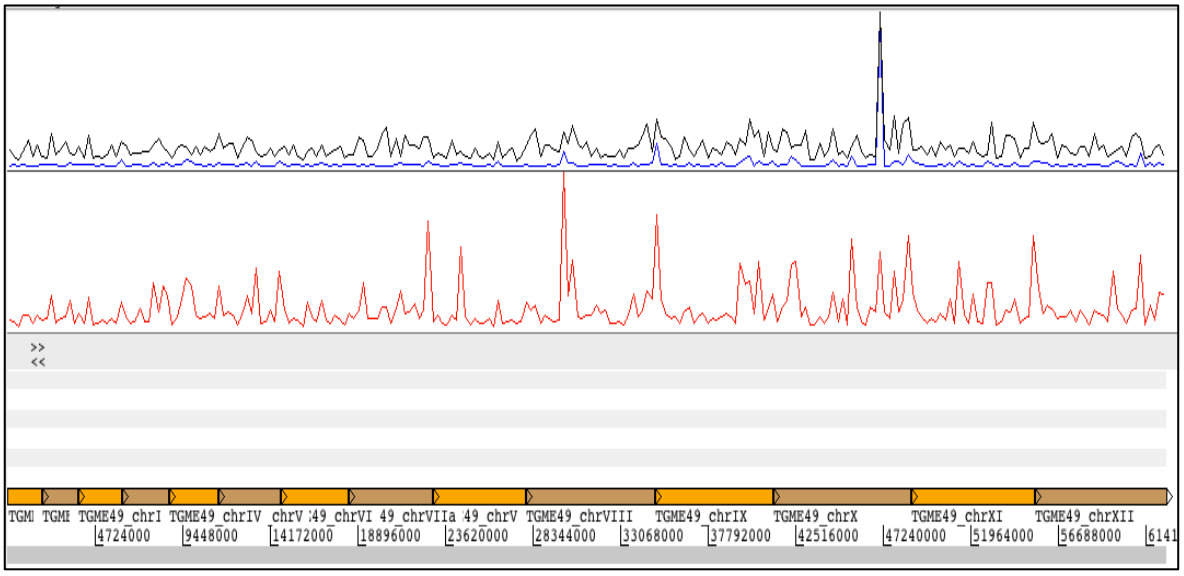

Figure 3.8: Coverage graphs of TgCkUg8 & 9 (mapped reads) to the reference genome chromosomes (TgME49) in blue and black respectively and SNPs graph in red (using Artemis)

It is noticed in Figures 3.6, 3.7 and 3.8; a peak in coverage graphs at one point in the genome for both Ugandan strains. By zooming this region using Artemis (version16.0.0), it is shown that there is an accumulation of repetitive reads in one region of the genome includes a hypothetical protein (TGME49_237120) and non-annotated region. There is therefore no obvious cause for this discrepancy (Figure 3.9).

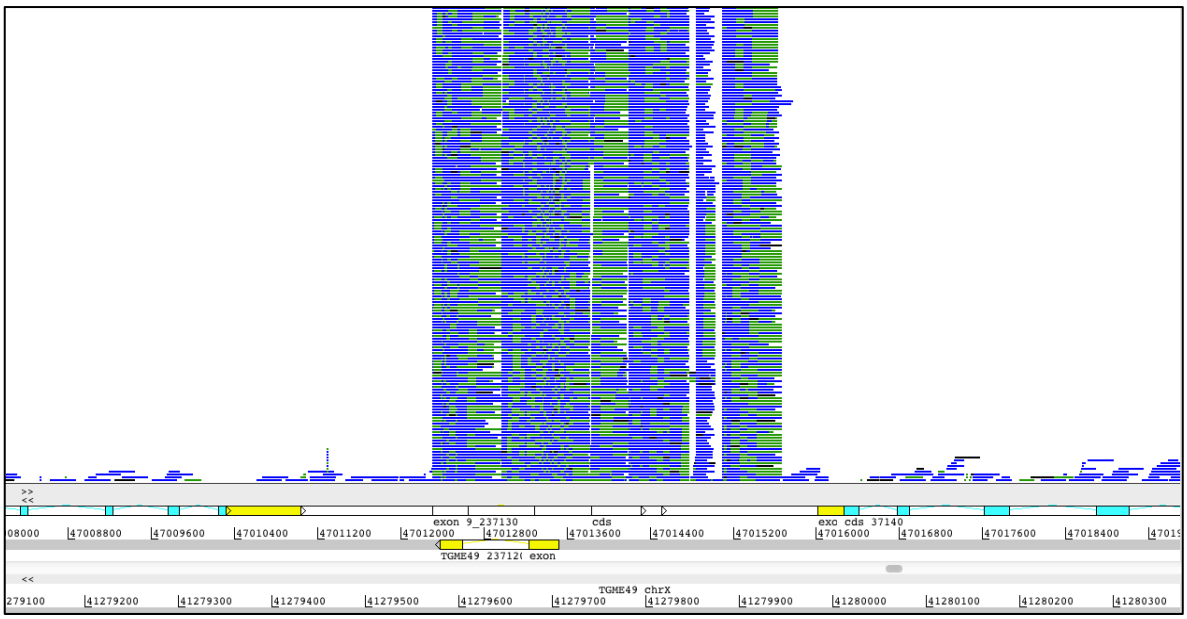

Figure 3.9: The peak in coverage graphs in the genome for TgCkUg8 and 9 strains (by use of Artemis).

In Figure 3.10, the distribution of SNPs in TgCkUg8 & 9 across 14 chromosomes of type II reference strains (TgME49) shows that the relative higher density of SNPs is located at the ends of most of chromosomes (telomeres), as explained above. It is noticeable that chromosome Ia has the lowest number of SNPs in both Ugandan strains compared with the other chromosomes.

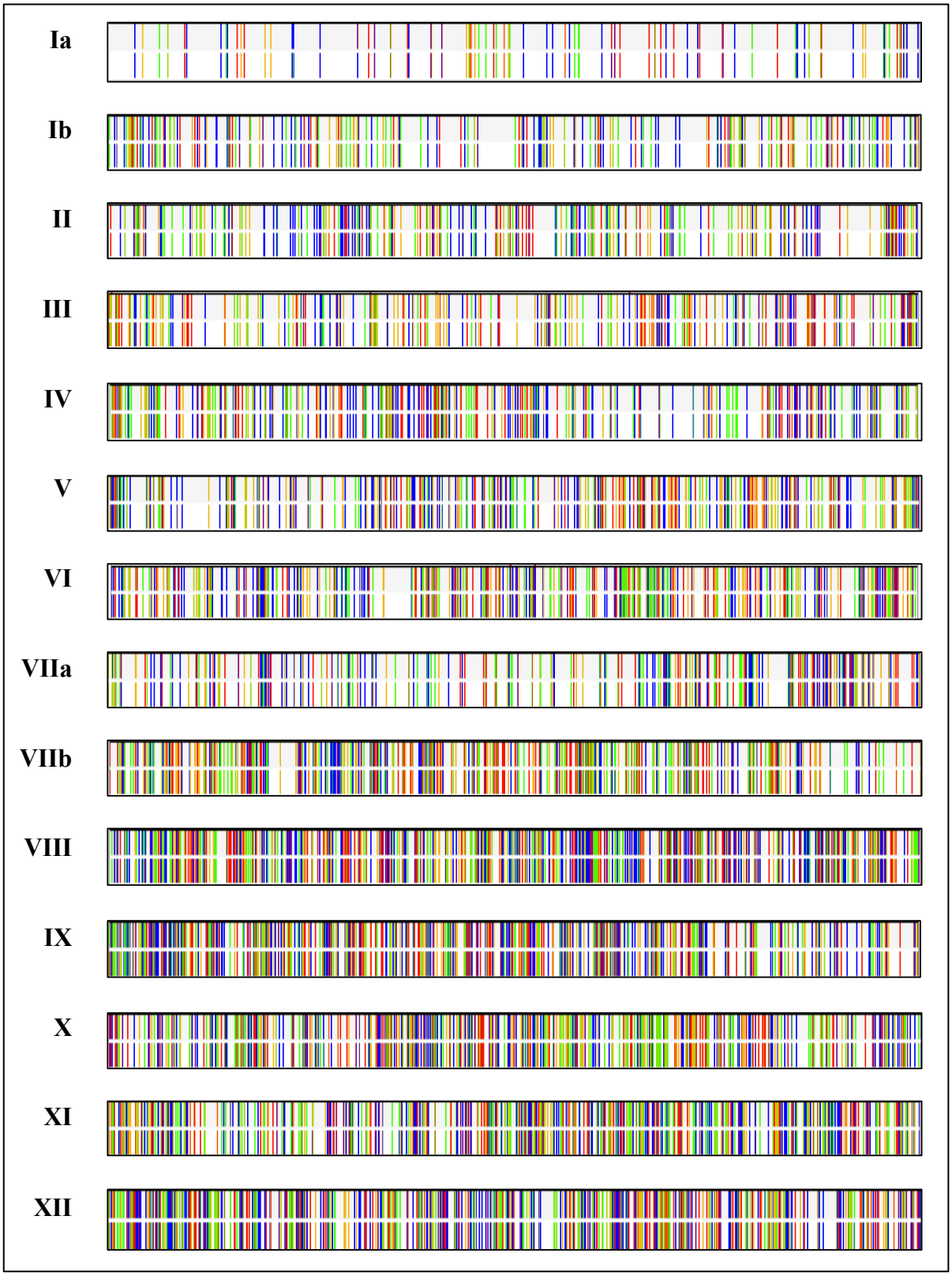

Figure 3.10: SNP distribution in TgCkUg8 (upper part) & 9 (lower part) across 14 chromosomes of type II reference strains (TgME49).
Colour indicates the different SNPs bases A, T, C, G. A similar colour in both TgCkUg8 (upper part) & 9 (lower part) indicates that they share the same SNPs i.e. the same base variation against the reference strains (TgME49) (using IGV genome browsers).

In comparison with TgCkUg2 (Bontell et al., 2009), the estimated SNPs/kb in all chromosomes of TgCkUg8 & 9 at 3x and 5x depths and TgCkUg2 at 4x depth compared with the reference genome TgME49 are summarised in Table 3.17.

Table 3.17: Summary of estimated SNPs/kb in all chromosomes of TgCkUg8 & 9 at 3x and 5x depths and TgCkUg2 at 4x depth compared with the reference genome TgME49

Chromosomes	SNPs/kb				
	TgCkUg8		TgCkUg9		TgCkUg2
	At 3x	At 5x	At 3x	At 5x	At 4x
Ia	0.5	0.6	0.5	0.4	0.1
Ib	1.5	3.3	0.7	**0.7**	0.9
II	1.3	9.1	0.8	**0.7**	**0.1**
III	2.3	7.5	1.1	1.2	2.1
IV	2.0	3.9	0.8	**0.8**	**0.1**
V	2.9	15.4	1.2	1.3	2.2
VI	1.9	8.9	0.6	**0.6**	**0.1**
VIIa	0.9	3.1	0.5	**0.5**	**0.1**
VIIb	1.1	1.8	0.6	**0.5**	1.3
VIII	1.2	2.9	0.6	**0.6**	1.2
IX	1.5	3.1	0.7	**0.7**	**0.1**
X	1.9	8.6	0.6	**0.6**	**0.1**
XI	1.6	7.9	0.4	0.4	0.1
XII	2.4	6.8	1.1	1.1	0.8
Total	**1.7**	**6.1**	**0.7**	**0.7**	**0.7**

In TgCkUg2, it is clear that type II chromosomes II, IV, VI, VIIa, IX and X (see Section 3.1.4) have the lowest estimated SNPs/kb, of 0.1, compared with type III chromosomes Ia, Ib, III, V, VIIb, VIII and XII (Note: chromosome Ia has the lowest estimated SNPs/kb in all Ugandan strains). By looking at the estimated rate of SNPs/kb in these chromosomes of TgCkUg9 at 5x depth, it is obvious that chromosomes Ib, II, IV, VI, VIIa, VIIb, VIII, IX and X have the lower estimated variation SNPs/kb, ranging from 0.5 to 0.8, compared with the other chromosomes. 6 of these 9 chromosomes II, IV, VI, VIIa, IX and X were considered as type II chromosomes in a Bontell et al., (2009) study, which suggested that these 9 chromosomes may be type II chromosomes in the TgCkUg9 strain.

The number of SNPs and estimated SNPs/kb in the genes, exons and introns for all chromosomes of TgCkUg8 and 9 at both 3x and 5x depths, compared with the reference genome TgME49, are shown in Tables 3.18, 3.19, 3.20 and 3.21.

Table 3.18: Number of SNPs and estimated SNPs/kb in the genes, exons and introns for all chromosomes of TgCkUg8 at 3x and 5x depths compared with the reference genome TgME49

Chromosomes	TgCkUg8											
	No. SNPs at 3x						No. SNPs at 5x					
	Genes	Exons	Introns	SNPs/ kb genes	SNPs/ kb exons	SNPs/ kb introns	Genes	Exons	Introns	SNPs/ kb genes	SNPs/ kb exons	SNPs/ kb introns
Ia	34	18	16	0.2	0.2	0.2	9	5	4	0.4	0.5	0.4
Ib	147	77	70	0.8	0.7	0.9	63	16	47	2.2	1.3	2.4
II	102	39	63	0.5	0.3	0.6	40	11	29	1.3	0.9	1.5
III	96	57	39	0.5	0.5	0.4	52	33	19	1.9	2.2	1.9
IV	153	106	47	0.6	0.7	0.5	59	31	28	1.8	1.8	2.8
V	161	86	75	0.5	0.5	0.5	80	49	31	1.9	2.6	1.6
VI	260	106	154	0.7	0.5	1.1	113	23	90	2.1	0.8	4.5
VIIa	212	52	160	0.5	0.2	0.9	156	23	133	2.6	0.8	4.4
VIIb	153	57	96	0.4	0.3	0.5	52	10	42	0.9	0.4	1.1
VIII	274	152	122	0.5	0.5	0.5	112	50	62	1.5	1.5	1.2
IX	342	198	144	0.5	0.5	0.5	188	105	83	1.9	2.2	2.1
X	255	135	120	0.4	0.3	0.4	119	45	74	1.3	0.9	1.9
XI	289	117	172	0.4	0.3	0.6	145	37	108	1.7	0.9	2.7
XII	630	279	351	0.9	0.7	1.1	401	130	271	3.9	2.8	4.5
Total (average)	**3108**	**1479**	**1629**	**0.5**	**0.4**	**0.6**	**1589**	**568**	**1021**	**1.8**	**1.4**	**2.5**

Table 3.19: Number of SNPs and estimated SNPs/kb in the genes, exons and introns for all chromosomes of TgCkUg9 at 3x and 5x depths, compared with the reference genome TgME49

Chromosomes	TgCkUg9											
	No. SNPs at 3x						No. SNPs at 5x					
	Genes	Exons	Introns	SNPs/ kb genes	SNPs/ kb exons	SNPs/ kb introns	Genes	Exons	Introns	SNPs/ kb genes	SNPs/ kb exons	SNPs/ kb introns
Ia	224	137	87	0.3	0.2	0.3	114	70	44	0.3	0.2	0.3
Ib	307	158	149	0.3	0.3	0.4	206	107	99	0.4	0.4	0.5
II	383	229	154	0.4	0.4	0.4	211	132	79	0.4	0.5	0.4
III	282	162	120	0.3	0.2	0.3	168	96	72	0.3	0.2	0.3
IV	395	215	180	0.3	0.3	0.4	239	116	123	0.4	0.3	0.5
V	523	299	224	0.4	0.3	0.4	305	170	135	0.4	0.4	0.5
VI	523	284	239	0.3	0.3	0.4	338	176	162	0.4	0.3	0.6
VIIa	553	255	298	0.2	0.2	0.4	414	160	254	0.3	0.2	0.5
VIIb	465	223	242	0.3	0.2	0.3	230	109	121	0.3	0.2	0.3
VIII	802	411	391	0.3	0.2	0.3	531	266	265	0.3	0.3	0.4
IX	943	556	387	0.3	0.3	0.3	610	348	262	0.3	0.3	0.4
X	1023	590	433	0.3	0.2	0.3	648	380	268	0.3	0.3	0.4
XI	755	451	304	0.2	0.2	0.2	489	282	207	0.3	0.2	0.3
XII	1330	697	633	0.4	0.3	0.5	931	463	468	0.5	0.4	0.7
Total (average)	**8508**	**4667**	**3841**	**0.3**	**0.3**	**0.3**	**5434**	**2875**	**2559**	**0.4**	**0.3**	**0.4**

It is clear from these data that the total estimated SNPs/kb in the introns is relatively higher than in exons of the genes in all chromosomes of both Ugandan strains. In TgCkUg8 genes, the total estimated SNPs/kb in introns is 0.6 compared with 0.4 in the exons at 3x depths, and 2.5 in introns, while it is 1.4 in exons at 5x depths (Table 3.20).

Table 3.20: Summary of estimated SNPs/kb in whole chromosomes, genes, exons and introns for all chromosomes of TgCkUg8 at 3x and 5x depths compared with the reference genome TgME49

Chromosomes	**TgCkUg8**							
	SNPs/kb at 3x				**SNPs/kb** at 5x			
	Total	Genes	Exons	Introns	Total	Genes	Exons	Introns
Ia	0.5	0.2	0.2	0.2	0.6	0.4	0.5	0.4
Ib	1.5	0.8	0.7	0.9	3.3	2.2	1.3	2.4
II	1.3	0.5	0.3	0.6	9.1	1.3	0.9	1.5
III	2.3	0.5	0.5	0.4	7.5	1.9	2.2	1.9
IV	2.0	0.6	0.7	0.5	3.9	1.8	1.8	2.8
V	2.9	0.5	0.5	0.5	15.4	1.9	2.6	1.6
VI	1.9	0.7	0.5	1.1	8.9	2.1	0.8	4.5
VIIa	0.9	0.5	0.2	0.9	3.1	2.6	0.8	4.4
VIIb	1.1	0.4	0.3	0.5	1.8	0.9	0.4	1.1
VIII	1.2	0.5	0.5	0.5	2.9	1.5	1.5	1.2
IX	1.5	0.5	0.5	0.5	3.1	1.9	2.2	2.1
X	1.9	0.4	0.3	0.4	8.6	1.3	0.9	1.9
XI	1.6	0.4	0.3	0.6	7.9	1.7	0.9	2.7
XII	2.4	0.9	0.7	1.1	6.8	3.9	2.8	4.5
Total (average)	**1.7**	**0.5**	**0.4**	**0.6**	**6.1**	**1.8**	**1.4**	**2.5**

Additionally, in TgCkUg9 genes, the total estimated SNPs/kb in introns and exons is similar by a value of 0.3 at 3x depth, while at 5x depth, it is higher in introns by 0.4 compared to exons at 0.3 SNPs/kb. This indicates greater variations within non-coding regions among the Ugandan strains compared to the reference strain TgME49 (Table 3.21).

Table 3.21: Summary of estimated SNPs/kb in whole chromosomes, genes, exons and introns for all chromosomes of TgCkUg9 at 3x and 5x depths compared with the reference genome TgME49

Chromosomes	TgCkUg9							
	SNPs/kb at 3x				**SNPs/kb** at 5x			
	Total	Genes	Exons	Introns	Total	Genes	Exons	Introns
Ia	0.5	0.3	0.2	0.3	0.4	0.3	0.2	0.3
Ib	0.7	0.3	0.3	0.4	0.7	0.4	0.4	0.5
II	0.8	0.4	0.4	0.4	0.7	0.4	0.5	0.4
III	1.1	0.3	0.2	0.3	1.2	0.3	0.2	0.3
IV	0.8	0.3	0.3	0.4	0.8	0.4	0.3	0.5
V	1.2	0.4	0.3	0.4	1.3	0.4	0.4	0.5
VI	0.6	0.3	0.3	0.4	0.6	0.4	0.3	0.6
VIIa	0.5	0.2	0.2	0.4	0.5	0.3	0.2	0.5
VIIb	0.6	0.3	0.2	0.3	0.5	0.3	0.2	0.3
VIII	0.6	0.3	0.2	0.3	0.6	0.3	0.3	0.4
IX	0.7	0.3	0.3	0.3	0.7	0.3	0.3	0.4
X	0.6	0.3	0.2	0.3	0.6	0.3	0.3	0.4
XI	0.4	0.2	0.2	0.2	0.4	0.3	0.2	0.3
XII	1.1	0.4	0.3	0.5	1.1	0.5	0.4	0.7
Total (average)	**0.7**	**0.3**	**0.3**	**0.3**	**0.7**	**0.4**	**0.3**	**0.4**

3.3.2 The level of local variation among type II strains

The available published type II strains genomes of *T. gondii* to estimate the level of local variations within type II strains are shown in Table 3.22.

Table 3.22: Type II strains of *Toxoplasma gondii* with whole genomic sequences in database (http://www.toxodb.org/toxo/)

NO.	Strains	Geographic	Host
1.	ME49	USA	Sheep
2.	PRU	France	Human
3.	ARI	USA	Human
4.	B41	USA	Bear
5.	B73	Unknown	Bear
6.	RAY	USA	Human

The variations among the type II Ugandan strains (TgCkUg8 & 9) in comparison with the reference type II genome (TgME49) can be seen with variations in the TgCkUg2 strain and other type II strains (Tables 3.23 and 3.24).

Table 3.23: Number of SNPs among type II and 12 strains in addition to type I (GT1) and III (VEG) reference genomes of *Toxoplasma gondii* against the type II reference genome (TgME49).
These data were retrieved from (http://www.toxodb.org/toxo/) through applying multiple queries for each chromosome between the type II strains by using the command: (Identify SNPs based on Genomic Location-chromosome)
(http://www.toxodb.org/toxo/showQuestion.do?questionFullName=SnpQuestions.NgsSnpsByLocation)

Chromosomes	**Chr. Length ($x10^5$)**	**No. of SNPs**						
		Type I reference	**Type III reference**	**Type II strains**		**Type 12 strains**		
		GT1	**VEG**	**B73**	**PRU**	**ARI**	**B41**	**RAY**
Ia	18.6	311	50	52	132	5721	7	5613
Ib	19.6	18434	7180	25	365	9621	31	3868
II	23.5	22187	22146	74	632	2744	1707	2057
III	25.3	23381	18894	103	535	2366	35	1949
IV	26.9	603	23846	75	524	8103	7949	7855
V	33.3	27768	28669	28550	988	14125	11434	11541
VI	36.6	32712	8964	8953	633	16114	71	8700
VIIa	45.4	33621	40864	40777	633	15304	13820	14077
VIIb	50.7	45515	25673	461	1042	3970	2980	3503
VIII	69.7	64142	30187	652	1223	11793	7662	8957
IX	63.3	53353	39859	667	1431	9360	88	6467
X	74.9	62431	62402	95	1455	27467	5481	19688
XI	66.2	56939	116	109	1204	8353	82	5591
XII	70.9	66198	22244	6643	1262	18950	2248	4896
TOTAL	**624.9**	**507595**	**331094**	**87236**	**12059**	**153991**	**53595**	**104762**

By comparing the estimated SNPs/kb rates in the Ugandan strains (TgCkUg2, 8 & 9) and 5 of type II strains (PRU, ARI, B41, B73 and RAY) against the type II reference strain (ME49), it is evident that both TgCkUg2 and 9 have the second lowest SNPs/kb rate at 0.7, while the lowest value is 0.2 in PRU strain. This indicates that these Ugandan strains (TgCkUg2 and 9) are closer to the type II reference strains compared with the other type II strains (except PRU).

Table 3.24: Estimated SNPs/kb among type II and 12 strains in addition to type I (GT1) and III (VEG) reference genomes of *Toxoplasma gondii* against with the type II reference genome (TgME49)

Chromosomes	Chr. Length ($x10^5$)	SNPs/kb									
		Type I reference	Type III reference	Type II strains					Type 12 strains		
		GT1	VEG	TgCkUg8 at 5x	TgCkUg9 at 5x	TgCkUg2 at 4x	B73	PRU	ARI	B41	RAY
Ia	18.6	0.2	0.03	0.6	0.4	0.1	0.03	0.1	3.1	0.004	3.02
Ib	19.6	9.4	3.7	3.3	0.7	0.9	0.01	0.2	4.9	0.02	1.9
II	23.5	9.5	9.4	9.1	0.7	0.1	0.03	0.3	1.2	0.7	0.9
III	25.3	9.2	7.5	7.5	1.2	2.1	0.04	0.2	0.9	0.01	0.8
IV	26.9	0.2	8.9	3.9	0.8	0.1	0.03	0.2	3.1	2.9	2.9
V	33.3	8.3	8.6	15.4	1.3	2.2	8.6	0.3	4.2	3.4	3.5
VI	36.6	8.9	2.5	8.9	0.6	0.1	2.4	0.2	4.4	0.02	2.4
VIIa	45.4	7.4	8.9	3.1	0.5	0.1	8.9	0.1	3.4	3.04	3.1
VIIb	50.7	8.9	5.1	1.8	0.5	1.3	0.1	0.2	0.8	0.6	0.7
VIII	69.7	9.2	4.3	2.9	0.6	1.2	0.1	0.2	1.7	1.1	1.3
IX	63.3	8.4	6.3	3.1	0.7	0.1	0.1	0.2	1.5	0.01	1.02
X	74.9	8.3	8.3	8.6	0.6	0.1	0.01	0.2	3.7	0.7	2.6
XI	66.2	8.6	0.02	7.9	0.4	0.1	0.02	0.2	1.3	0.01	0.8
XII	70.9	9.3	3.1	6.8	1.1	0.8	0.9	0.2	2.7	0.3	0.7
TOTAL (Average)	**624.9**	**8.1**	**5.3**	**6.1**	**0.7**	**0.7**	**1.4**	**0.2**	**2.5**	**0.9**	**1.7**

3.4 DISCUSSION

Whole genome sequencing of two Ugandan strains (TgCkUg8 and 9) through paired-end Illumina sequencing method revealed variations between these strains and against the type II reference strain of *T. gondii* (TgME49). The coverage percentage of the type II reference strain genome of *T. gondii* - TgME49 (62.5 Mbp) - was low in TgCkUg8 sequences with only 12.0% and 1.7% at 3x and 5x depths respectively. In addition, this percentage was relatively low in TgCkUg9 with a total coverage of 57.15% and 30.4% at 3x and 5x depths respectively. These estimated total coverage percentages are considered to be low in comparison with the total coverage of sequences generated by the 454 GSFLX method for TgCkUg2 which was 84.09% of the reference genome TgME49 (61.6 Mbp) at 4x depth, even though SNPs were observed across all 14 chromosomes of the reference strains (TgME49) for both Ugandan strains. These SNPs in both Ugandan strains were called at 3 reads depth, with a total number of SNPs of 12574 in TgCkUg8 and 24643 in TgCkUg9. To increase the confidence in SNPs, they were called at 5 reads depth and distributed across all 14 chromosomes of the reference strains (TgME49), with a total number of 6593 in TgCkUg8 compared to 13524 in TgCkUg9. This variation in the number of SNPs in TgCkUg8 and 9 at both 3 and 5 reads depths may be explained by the variation in the estimated total coverage percentages in these Ugandan sequences. This correlation between the coverage percentage of the Ugandan strains sequences and the number of SNPs in these strains is clearly shown in Figures 3.6 and 3.7.

To check the reliability of TgCkUg8 and 9 strain sequences, a comparison through alignment has been applied between the MiSeq sequences for both strains with the Sanger sequence data for specific loci generated in chapter 2, and the Sanger sequence data for 34 loci generated by Bontell et al., (2009) (Appendix 3). All the covered MiSeq sequences of TgCkUg8 and 9 for these loci were identical for particular strain. This confirms the reliability of TgCkUg8 and 9 strain sequences.

In TgCkUg8, it is evident that the estimated SNP/kb rate is higher at 5x depth compared to 3x (Table 3.15). This variation in SNP/kb rate could be explained by the great variation in the coverage for TgCkUg8 at 3x and 5x depths against the reference genome TgME49 (Figure 3.11). Furthermore, it can be seen from Table 3.15 that the total number of SNPs in TgCkUg8 is decreased by about the half from 12574 at 3x to 6593 at 5x depth. However, the number of bases covered falls by about the seventh from 74.7 $X10^5$ at 3x to 10.9 $X10^5$ at 5x

depth. This obvious variation in the coverage resulted in obvious difference when calculating SNPs/kb at both depths.

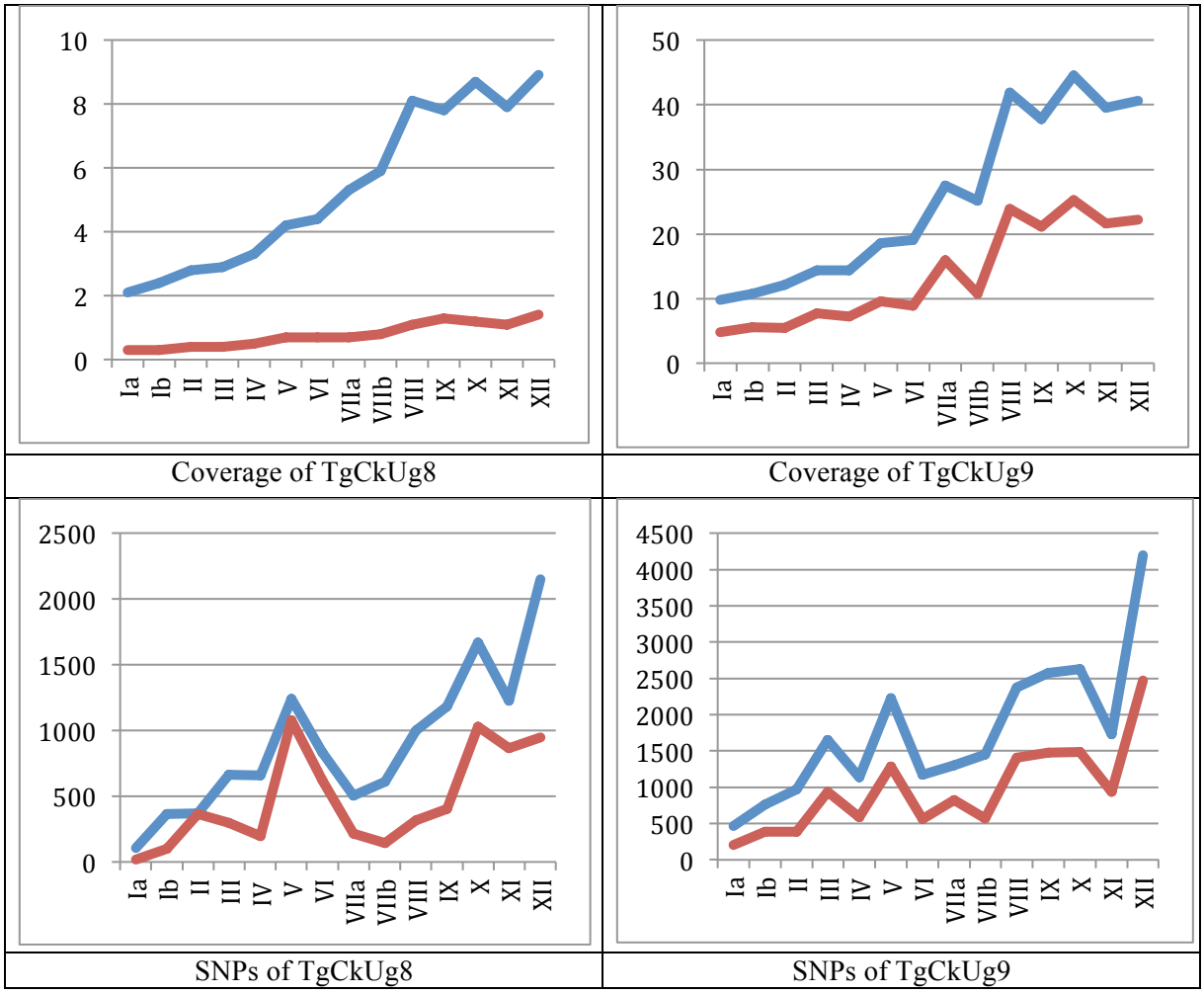

Figure 3.11: Coverage and SNPs graphs of TgCkUg8 & 9 at both 3x (blue) and 5x (red) read depths.

The identification of SNPs in TgCkUg8 and 9 against the type II reference genome (TgME49) indicated that these Ugandan strains are not identical to the reference strains. In addition, detection of SNPs that had been shared between TgCkUg8 and 9 against the type II reference genome of *Toxoplasma gondii* TgME49 with 6385 SNPs at 3 reads depth, and 3251 at 5 reads depth, indicates that the variations between these Ugandan strains are lower than variations with the type II reference strains (TgME49).

By observing the distribution of SNPs across the 14 chromosomes of the reference strains for both Ugandan strains in Figures 3.5, 3.6, 3.7 and 3.8, it could be shown that the density of SNPs is higher at the ends of most of chromosomes (telomeres) compared with the central

region of the chromosome. This pattern of telomeric diversity, which included several of the surface antigens and central genetic conservation, had been reported in many protozoan pathogens, such as *Plasmodium falciparum* (Gardner et al., 2002); *Trypanosoma brucei* (Berriman et al., 2005); *Trypanosoma cruzi* (El-Sayed et al., 2005); and *Leishmania major* (Ivens et al., 2005). In *Toxoplasma gondii*, the surface antigens genes are located in the telomeres of most of the chromosomes (Jung et al., 2004).

It is noticed in Figure 3.5 that the pattern of SNPs at the telomeres varies between the 14 chromosomes, with very high rates at the telomeres of 5 chromosomes (III, IV, V, IX, XII). By using IGV genome browser, it is shown that most of these telomeric SNPs located within non annotated regions within these 5 chromosomes, in addition to two regions annotated as hypothetical proteins in chromosome XII (TGME49_300790 and TGME49_200110).

By comparing the density of SNPs distributed across all 14 chromosomes of the reference strains (TgME49) for both TgCkUg8 and 9 sequences in Tables 3.14, 3.15 and 3.16**,** and Figures 3.5, 3.6, 3.7, 3.8 and 3.10, it is noticeable that chromosome Ia has the lowest number of SNPs in both Ugandan strains. This pattern of low diversity in the shortest chromosome Ia is also clearly shown in TgCkUg2 (Table 3.17**)** (Bontell et al., 2009). This finding is in agreement with a study by Khan et al. (2006a) which revealed that chromosome Ia had the lowest diversity compared with other chromosomes in the type I reference strain (RH) and type II reference strain (ME49). Khan activated the whole genome sequencing of RH strain genome via the Sanger sequencing technique and compared it with the published whole genome sequence of TgME49 strain genome.

In TgCkUg9, by observing the estimated SNPs/kb in all chromosomes of 5x depth it could be shown that chromosomes Ib, II, IV, VI, VIIa, VIIb, VIII, IX and X have the lowest estimated SNPs/kb, ranging from 0.5 to 0.8 (Note: chromosome Ia has the lowest estimated SNPs/kb in all Ugandan strains). In TgCkUg2, 6 of these 9 chromosomes II, IV, VI, VIIa, IX and X have the lowest estimated SNPs/kb, of 0.1, compared with other chromosomes, and are considered as type II chromosomes by Bontell et al. (2009) (see Section 3.1.4). This suggests that these 9 chromosomes originate from a type II parent strain in TgCkUg9 strain.

The level of local variations within type II strains was estimated through calculation of SNPs/kb rates within the Ugandan strains (TgCkUg2, 8 & 9) and 5 of the available type II strain genomes in the database against the type II reference strain (ME49). It was shown that

both TgCkUg2 and 9 have the second lowest SNPs/kb rate at 0.7, which indicates that these Ugandan strains are more similar to the type II reference strain than the other type II strains (except PRU strains which are closer to the reference strain). The relatively high SNPs/kb rate of 6.1 in TgCkUg8 strain can be explained due to the very low coverage of the sequences for this strain at x5 reads depth.

4. CHAPTER 4: SNP discovery in Key Biologically Relevant Gene Families

4.1. INTRODUCTION

A number of gene families have been well-researched, as they have important biological functions. These families include the surface (SRS), rhoptry (ROPs) and dense granule (GRA) proteins. In this Chapter, we look specifically at variation in these genes in relation to their biological functions.

The tachyzoite of *T. gondii* is crescent shaped with an anterior pointed end (apical complex) and a posterior rounded end. The apical complex is involved in attachment, invasion and formation of the parasitophorous vacuole membrane (PVM) (Kim and Weiss, 2004). This apical complex contains two groups of secretory organelles called micronemes and rhoptries. The micronemes organelles are involved in facilitating the attachment and invasion of the host cell (Kim and Weiss, 2004). After attachment, these rhoptries release their contents at the beginning of invasion. The rhoptry organelles are club-shaped membranous sacs which open onto the anterior of the parasite. They contain two groups of proteins - those located in the neck of the parasite, called rhoptry neck proteins or RONs, and those found in the posterior bulb, called rhoptry bulb proteins or ROPs (Bradley et al., 2005). The parasite also contains dense granules which are more widely distributed towards anterior and posterior ends.

During the process of invasion, the surface and secretory proteins (rhoptries and dense granules) are essential and perform important roles. Various members within these 3 protein families have attracted the interest of scientists recently as they are shown to be essential for the differences in mouse virulence that appear within different strains of *T. gondii*. SAG (surface antigen) proteins of *T. gondii* have an important role in the initial stage of attachment to the host cell (Grimwood and Smith, 1992; Robinson et al., 2004). It has also been suggested that variability within this family of surface proteins facilitates the process of the invasion of the host cell (Grimwood and Smith, 1996; Jung et al., 2004). The process of the parasite's invasion of the host cell is introduced by contact between the apical complex of the parasite and the surface of the host cell which is followed by internalisation of the parasite at the site of contact, leading to formation of PVM, then closing of the vacuole

behind the parasite (Dubremetz et al., 1998). During this invasion process, exocytosis of contents of both rhoptry proteins groups (ROPs and RONs), which contain lipid and proteins, occurs in the PVM. The rhoptry proteins are suggested to be essential in the formation of PVM. Many ROPs migrate either to the parasitophorous vacuole membrane or into the cytosol of the host cell and even the nucleus, where they interfere with the gene expression of the host cell (Boothroyd and Dubremetz, 2008). After closure of the PVM, the dense granules start to exocytose their proteins in the vacuole which continue to be released throughout the intracellular development of the parasite in the intermediate host (Dubremetz et al., 1998). These dense granule proteins interact with several host proteins and have an important role in the persistence of infection and the inhibition of apoptosis of the host cell (Coppens et al., 2006).

The ROP proteins may also play a significant role in the modulation of inflammation during parasite infection depending on the ROP protein involved, resulting in a remarkable variation in virulence between different strains. Genetic variation between members of the ROP family was shown to be a significant factor in disease pathogenesis. This can be through two different mechanisms: firstly, by activation of the host immune system, leading to severe inflammation-related pathology; secondly, cellular damage, driven via unrestricted proliferation of *T. gondii* (Melo et al., 2011).

4.1.1. Surface proteins (SAG1- related sequences) (SRS)

In *Toxoplasma gondii*, many surface antigens have been identified. These glycophosphatidyl inositol (GPI) anchored proteins are members of a family of proteins defined by the first member to be described, SAG1 or p30 (Burg et al., 1988; Kasper et al., 1983). These SAG1-related sequences (SRS) proteins are encoded by about 100 genes (Jung et al., 2004) (Table 4.1).

The relatively large number of SRS proteins that are differentially expressed by the developmental stages of the parasite could imply variation in their function (Nagel and Boothroyd, 1989). Some members of the SRS family are stage-specific proteins. This group of antigens includes the tachyzoites-specific surface antigens such as SAG1 and SAG2, which are highly immunogenic (Grimwood and Smith, 1996), and bradyzoites-specific antigens such as SRS9 and BAG1 - bradyzoite antigen 1 (Kim and Boothroyd, 2005). There is evidence that some of these proteins are involved in adhesion at the initial host cell

attachment, whereas others may influence the immune response by attracting this response and redirecting it away from more important membranes. It was revealed that the strong immune response induced by tachyzoite-specific SRS antigens may have a role in the persistence of parasite infection by reducing the immune response against bradyzoite SRS antigens (Kim and Boothroyd, 2005).

SAG1 (P30) is a GPI anchored surface protein that was identified in the early molecular studies of *T. gondii* .It has the ability to produce a strong antibody response in humans and mice. Its significance in the parasite lifecycles is worthy of further invesigation as it is the prototype of a large family of surface proteins that have a role in parasite virulence and are expressed differently during the lifecycle (Lekutis et al., 2001). It has been suggested that this surface protein is involved in attachment to the host cell before invasion and modulates the host's immune system (Boothroyd et al., 1998).

SAG3 is another member of the surface protein family that may be involved directly in parasite virulence. It has been shown that SAG3 knockout parasites have lower attachment-invasion rates *in vitro* and decreased lethality (Dzierszinski et al., 2000).

Table 4.1: SAG1- related sequence protein (SRS) genes of *Toxoplasma gondii* ME49 reference strain.
Data retrieved from TOXODB by applying multiple queries by using the tool: (Identify Genes based on Text). NA: Not Assigned.
(http://www.toxodb.org/toxo/showQuestion.do?questionFullName=GeneQuestions.GenesByTextSearch)

NO.	Gene ID	Chr.	(SRS)	NO.	Gene ID	Chr.	(SRS)	NO.	Gene ID	Chr.	(SRS)
1.	TGME49_295600	Ia	SRS10	38.	TGME49_259290	VIIb	SRS26C	75.	TGME49_264660	IX	SRS44
2.	TGME49_208850	Ib	SRS11 (SAG2B)	39.	TGME49_259270	VIIb	SRS26E	76.	TGME49_235600	X	SRS45
3.	TGME49_321490	Ib	SRS12A	40.	TGME49_329710	NA	SRS26I	77.	TGME49_214190	X	SRS46
4.	TGME49_321480	Ib	SRS12B	41.	TGME49_329700	NA	SRS26J	78.	TGME49_275355	X	SRS47A
5.	TGME49_321470	Ib	SRS12D	42.	TGME49_258810	VIIb	SRS27B	79.	TGME49_275360	X	SRS47B
6.	TGME49_222370	II	SRS13	43.	TGME49_258550	VIIb	SRS28	80.	TGME49_275370	X	SRS47C
7.	TGME49_254060	III	SRS14A	44.	TGME49_233450	VIII	SRS29A (SRS1)	81.	TGME49_275380	X	SRS47D
8.	TGME49_320250	IV	SRS15A	45.	TGME49_233460	VIII	SRS29B (SAG1)	82.	TGME49_296640	X	SRS48E
9.	TGME49_320240	IV	SRS15B	46.	TGME49_233480	VIII	SRS29C (SRS2)	83.	TGME49_207010	X	SRS48K
10.	TGME49_320230	IV	SRS15C	47.	TGME49_273130	VIII	SRS30A	84.	TGME49_207005	X	SRS48Q
11.	TGME49_320200	IV	SRS16A	48.	TGME49_273120	VIII	SRS30C	85.	TGME49_207130	X	SRS49A (SAG2Y)
12.	TGME49_320190	IV	SRS16B (SRS9)	49.	TGME49_273110	VIII	SRS30D	86.	TGME49_207140	X	SRS49B (SAG2X)
13.	TGME49_320180	IV	SRS16C (BSR4)	50.	TGME49_272560	VIII	SRS31B	87.	TGME49_207150	X	SRS49C (SAG2D)
14.	TGME49_320170	IV	SRS16E (SRS6)	51.	TGME49_271980	VIII	SRS32	88.	TGME49_207160	X	SRS49D (SAG2C)
15.	TGME49_319360	IV	SRS17A	52.	TGME49_243790	VI	SRS33	89.	TGME49_311440	XI	SRS50
16.	TGME49_319350	IV	SRS17B	53.	TGME49_271050	VIII	SRS34A (SAG2A)	90.	TGME49_308840	XI	SRS51 (SRS3)
17.	TGME49_215680	X	SRS18 (SAG2E)	54.	TGME49_280570	VIIa	SRS35A	91.	TGME49_315320	XI	SRS52A
18.	TGME49_301140	IV	SRS19A	55.	TGME49_280580	VIIa	SRS35B	92.	TGME49_315340	XI	SRS52C
19.	TGME49_301150	IV	SRS19B (SRS8)	56.	TGME49_292250	IX	SRS36A	93.	TGME49_315330	XI	SRS52D
20.	TGME49_301160	IV	SRS19C (SRS7)	57.	TGME49_292260	IX	SRS36B (SAG5D)	94.	TGME49_315345	XI	SRS52F
21.	TGME49_301170	IV	SRS19D	58.	TGME49_292270	IX	SRS36C (SAG5A)	95.	TGME49_315370	XI	SRS53A
22.	TGME49_301180	IV	SRS19F	59.	TGME49_292280	IX	SRS36D (SAG5C)	96.	TGME49_315380	XI	SRS53B
23.	TGME49_285870	V	SRS20A	60.	TGME49_292275	IX	SRS36E	97.	TGME49_315390	XI	SRS53C
24.	TGME49_285860	V	SRS20C	61.	TGME49_210320	IX	SRS37A	98.	TGME49_315400	XI	SRS53D
25.	TGME49_283460	V	SRS21	62.	TGME49_210330	IX	SRS37B	99.	TGME49_315410	XI	SRS53F
26.	TGME49_238440	VI	SRS22A	63.	TGME49_267130	IX	SRS38A	100.	TGME49_315740	XI	SRS54
27.	TGME49_238460	VI	SRS22B	64.	TGME49_267140	IX	SRS38B	101.	TGME49_309300	XI	SRS55A
28.	TGME49_238470	VI	SRS22C	65.	TGME49_267150	IX	SRS38C	102.	TGME49_309330	XI	SRS55F
29.	TGME49_238480	VI	SRS22D	66.	TGME49_267160	IX	SRS38D	103.	TGME49_219348	XI	SRS55M
30.	TGME49_238490	VI	SRS22E	67.	TGME49_281930	VIIa	SRS39	104.	TGME49_309352	XI	SRS55N
31.	TGME49_238500	VI	SRS22F	68.	TGME49_224790	X	SRS40A	105.	TGME49_246070	XII	SRS56
32.	TGME49_238520	VI	SRS22G	69.	TGME49_224780	X	SRS40B	106.	TGME49_308020	XII	SRS57 (SAG3)
33.	TGME49_238850	VI	SRS22I	70.	TGME49_224770	X	SRS40D	107.	TGME49_251958	XII	SRS59B
34.	TGME49_239090	VI	SRS23	71.	TGME49_224760	X	SRS40E (SRS4)	108.	TGME49_279000	XII	SRS59J
35.	TGME49_213280	V	SRS25	72.	TGME49_224750	X	SRS40F	109.	TGME49_251962	XII	SRS59K
36.	TGME49_259410	VIIb	SRS26A	73.	TGME49_234370	X	SRS42	110.	TGME49_224170	X	SRS60A
37.	TGME49_259300	VIIb	SRS26B	74.	TGME49_234930	X	SRS43	111.	TGME49_226860	X	SRS67

4.1.2. Rhoptry proteins (ROPs)

ROP proteins are a diverse set of molecules that are stored in the rhoptry organelles and secreted into the parasitophorous vacuole membrane (PVM) to mediate the intracellular interactions with the host cell. ROP2 was the first member identified in this family (Sadak et al., 1988), following genomic analysis. This has now grown to around 40 candidate proteins (Table 4.2) (Peixoto et al., 2010). Although, the first ROP proteins were identified almost 20 years ago, their functional role remained uncertain until recent genomic and proteomic studies revealed that some members of this family were true kinases and that they were translocated, during the invasion, in the host cell which increased interest in further investigations (El Hajj et al., 2006; Saeij et al., 2006).

Genetic studies of *T. gondii* virulence in mice narrowed the responsible chromosomal loci to certain genes, including those that encode members of this family (Taylor et al., 2006). At least three ROP family members, ROP16, 18 and 5, were detected by mapping the major virulence genes in progeny and genetic crosses in mice, indicating that *T. gondii* genetic studies in mice often identify variant rhoptry as candidate virulence genes (Saeij et al., 2006; Taylor et al., 2006). The kinases ROP16 and ROP38 were shown to play a role in moderating inflammation (Ong et al., 2010; Peixoto et al., 2010; Yamamoto et al., 2009). The kinase ROP18 has a function in protecting the parasitophorous vacuole membrane (PVM) against destruction (Fentress et al., 2010; Steinfeldt et al., 2010). This phenotype requires the expression of the virulence-associated pseudokinase ROP5 (Niedelman et al., 2012). It has been shown that pseudokinases play a significant role in the regulation of active kinases (Boudeau et al., 2006). In addition, the discovery of these rhoptry proteins and identification of a rhoptry phosphatase translocation in the host cell provides evidence for a phenomenon which had long been hypothesized that the parasite during invasion has the ability to translocate proteins directly into host cell cytosol (Gilbert et al., 2007). Furthermore, a member of the ROPs proteins called toxofilin was shown to be involved in the enhancement of host cell invasion through disruption of the cortical actin skeleton in the host cell throughout the process of invasion (Delorme-Walker et al., 2012).

It has also been shown that some of the rhoptry neck proteins (RONs) have a significant role in host cell invasion through the formation of the moving junction, a structure in Apicomplexa that facilitates the invasion of host cells (El Hajj et al., 2006; Sadak et al., 1988). Structurally, the moving junction is necessary for host cell invasion, but only subsets

of RONs (RON 2, 4, 5 and 8) are considered to be essential to its formation (Besteiro et al., 2011). Deletion of a member of these proteins called RON8 does not disturb the development of the parasite, nor cause any decline of virulence in mice via the reduction of parasite invasion to host cells (El Hajj et al., 2007).

Table 4.2: Rhoptry protein (ROPs) genes of *Toxoplasma gondii* ME49 reference strain.
Data retrieved from TOXODB by applying multiple queries by using the tool: (Identify Genes based on Text). NA: Not Assigned
(http://www.toxodb.org/toxo/showQuestion.do?questionFullName=GeneQuestions.GenesByTextSearch)

NO.	Gene ID	CHr.	Rhoptry proteins (ROPs)	NO.	Gene ID	CHr.	Rhoptry proteins (ROPs)
1.	TGME49_294560	Ia	ROP37	32.	TGME49_263220	VIIb	ROP21
2.	TGME49_295105	Ia	ROP, putative	33.	TGME49_270920	VIII	ROP32
3.	TGME49_295110	Ia	ROP7	34.	TGME49_230470	VIII	ROP46, putative
4.	TGME49_295125	Ia	ROP4	35.	TGME49_229010	VIII	RON4
5.	TGME49_207610	Ib	ROP36	36.	TGME49_291960	IX	ROP40
6.	TGME49_207700	Ib	ROP22	37.	TGME49_266100	IX	ROP41
7.	TGME49_297960	II	RON6	38.	TGME49_306060	IX	RON8
8.	TGME49_252360	III	ROP24	39.	TGME49_265120	IX	RON, putative
9.	TGME49_253330	III	ROP	40.	TGME49_227810	X	ROP11
10.	TGME49_211260	IV	ROP26	41.	TGME49_227010	X	ROP30
11.	TGME49_211290	IV	ROP15	42.	TGME49_215775	X	ROP8
12.	TGME49_239600	VI	ROP23	43.	TGME49_223920	X	RON3
13.	TGME49_240090	VI	ROP34, putative	44.	TGME49_215785	X	ROP2A
14.	TGME49_242110	VI	ROP38	45.	TGME49_214080	X	toxofilin
15.	TGME49_242230	VI	ROP29	46.	TGME49_314250	XI	bradyzoite rhoptry protein BRP1
16.	TGME49_242240	VI	ROP19A	47.	TGME49_313330	XI	ROP27
17.	TGME49_242250	VI	ROP19B	48.	TGME49_315210	XI	ROP, putative
18.	TGME49_243730	VI	ROP9	49.	TGME49_315940	XI	ROP, putative
19.	TGME49_201130	VIIa	ROP33	50.	TGME49_310010	XI	RON1
20.	TGME49_202780	VIIa	ROP25	51.	TGME49_311470	XI	RON5
21.	TGME49_203990	VIIa	ROP12	52.	TGME49_308810	XI	RON9
22.	TGME49_205250	VIIa	ROP18	53.	TGME49_309590	XI	ROP1
23.	TGME49_304740	VIIa	ROP35	54.	TGME49_315490	XI	ROP10
24.	TGME49_258230	VIIb	ROP20	55.	TGME49_312270	XI	ROP13
25.	TGME49_258370	VIIb	ROP28	56.	TGME49_315220	XI	ROP14
26.	TGME49_258580	VIIb	ROP17	57.	TGME49_308093	XII	ROP
27.	TGME49_258660	VIIb	ROP6	58.	TGME49_308096	XII	ROP
28.	TGME49_258800	VIIb	ROP31	59.	TGME49_249470	XII	ROP
29.	TGME49_261750	VIIb	RON10	60.	TGME49_300100	XII	RON2
30.	TGME49_262050	VIIb	ROP39	61.	TGME49_308090	XII	ROP5
31.	TGME49_262730	VIIb	ROP16	62.	TGME49_327200	NA	RON, putative

4.1.3 Dense granule proteins (GRA)

In *T. gondii* dense granules, about twenty proteins were identified and shown to be exocytosed in the vacuole during or after invasion of the host cell (Table 4.3). It is suggested that the only protein which is essential to parasite survival is GRA1, as its coding gene cannot be deleted without causing lethal mutation. GRA2 and GRA3 deletion results in a significant reduction in parasite virulence in mice; however, only minor changes occur during *in vitro* development or in observation of small morphological modifications (Craver and Knoll, 2007; Mercier et al., 1998). On the other hand, parasite virulence in mice was not changed by deletion of GRA5 or GRA14 (Mercier et al., 2001; Rome et al., 2008). GRA7 is shown to be essential for parasites in low nutrition status during *in vitro* development, but there is no information on its role *in vivo* where the occurrence of starvation is unlikely. It thus cannot be considered as a virulence factor.

Recently, the protein GRA15 was reported to have an effect on the host immune defence. In *Toxoplasma gondii* strains, GRA15 displays genetic variation and in type II strains has a significant role in the modulation of the host immune response through activation of NFKB signaling greater than type I and III strains (Rosowski et al., 2011).

Table 4.3: Dense granular protein (GRA) genes of *Toxoplasma gondii* ME49 reference strain. Data were retrieved from TOXODB by applying multiple queries by using the tool: (Identify Genes based on Text) (http://www.toxodb.org/toxo/showQuestion.do?questionFullName=GeneQuestions.GenesByTextSearch)

NO.	Gene ID	CHr.	Dense granular proteins (GRA)	NO.	Gene ID	CHr.	Dense granular proteins (GRA)
1.	TGME49_222170	II	DG32	10.	TGME49_268900	VIII	GRA10
2.	TGME49_297880	II	DG32	11.	TGME49_270250	VIII	GRA1
3.	TGME49_275850	III	GRA12	12.	TGME49_288650	IX	GRA12
4.	TGME49_254720	III	GRA8	13.	TGME49_275470	X	GRA15
5.	TGME49_212410	IV	GRA11	14.	TGME49_227620	X	GRA2
6.	TGME49_237800	IV	GRA11	15.	TGME49_227280	X	GRA3
7.	TGME49_239740	VI	GRA14	16.	TGME49_275440	X	GRA6
8.	TGME49_286450	V	GRA5	17.	TGME49_310780	XI	GRA4
9.	TGME49_203310	VIIa	GRA7	18.	TGME49_251540	XII	GRA9

4.1.4 Virulence association with key biologically relevant gene families in *Toxoplasma gondii*

In *T. gondii*, one definition of virulence is the number of tachyzoites required to kill a mouse after intraperitonial injection. According to this definition, virulence can extend from complete virulence in which infection by a single tachyzoite results in death, to avirulence where there is an inability of any dose to kill the mouse. This definition remains accepted widely, and was particularly useful in investigating *Toxoplasma gondii*-mouse interactions. However, by studying other hosts, especially humans, the issue becomes less clear. In this case, determination of progression in pathology related to septicaemia or to organ localization (brain, eye, lung) is required. However, currently epidemiological data is only associated with parasite strain variation and the relationship between pathology and virulence factors has not yet been recognized (Boothroyd and Grigg, 2002). The ability of *T. gondii* to infect a wide range of hosts complicates the definition of virulence in this parasite. Among this broad host range, the variability of susceptibility to acute disease and to infection is obvious, and it is likely that both host and parasite factors contribute to these patterns of virulence.

T. gondii must exist inside host cells to be able to complete the lifecycle, as it is an obligatory parasite. Thus, the definition of virulence in this parasite is not based on the capability of the parasite to infect the host, but depends on the progression of the disease. Therefore, although the necessary gene products for host invasion are essential for parasite survival, they will not necessarily be virulence factors (Pernas and Boothroyd, 2010). Therefore, the definition of virulence factors should be based on gene products which are involved in and affect the severity of the disease, rather than those necessary for survival and transmission of the parasite. Virulence involves factors that have the ability to change the host cell in various ways, resulting in the enhancement of parasite proliferation, modifications to host immune systems, or improvement of intracellular resistance of the parasite against innate immune response.

The mouse model was extensively used to analyse the basis of the virulence of *T. gondii* and much was learned about parasite virulence factors in this system. This does not necessarily translate to virulence in other hosts. In mice, the stimulation of host immune response and the multiplication rate of *T. gondii* are considered to be the main factors in determining the parasite virulence, depending on the strain of the mouse used and the strain of *T. gondii*

(Gazzinelli et al., 1994). Recently, it became possible to have more insight into this model due to two factors. Firstly, through achievements in parasite genomics which have led to analysis and determination of the role of specific genes during infection; secondly, through the study and analysis of the host cell response to infection via transcriptomic analysis (Dubremetz and Lebrun, 2012).

Although virulence is complex and involves multiple pathways, there are key biologically relevant groups of genes that are involved in mediating the parasite's intracellular lifestyle. These groups include surface (SRS), rhoptry (ROPs) and dense granules (GRA) proteins. In many studies, some members of these three families were associated with virulence in mice. Initially, a genetic linkage analysis of different six RFLP markers in five *T. gondii* strains revealed that parasite virulence in mice was linked to a surface gene (SAG1) located on chromosome VIII (Howe et al., 1996). Experimental crosses between different strains of *T. gondii* with different virulence were considered as a valuable approach to the screening and detection of virulence-associated loci (Saeij et al., 2006; Taylor et al., 2006). The first detection of a major virulence-associated locus located on chromosome VIIa was achieved via linkage analysis of a recombinant progeny resulting from an experimental cross (Su et al., 2002). Four virulence-associated loci located on chromosomes VIIa (ROP18), VIIb (ROP16) and XII (SAG3 and ROP5), were identified through an experimental cross between type II and III strains of *T. gondii* (Saeij et al., 2006).

4.1.5 Aims

The aims of this study were to analyse the whole genome sequence of two sympatric *T. gondii* (type II) strains (TgCkUg8 and 9) and to compare them with the type II reference strain genome (TgME49); to seek novel allelic variants within three biologically relevant gene families for assessment of local strain diversity; and to look specifically at variation in these genes in relation to their biological functions.

4.2. METHODS

4.2.1. Identification of genes of three biologically relevant gene families

The genes of three biologically important families of proteins (SAG1- related sequences (SRS) genes, rhoptries (ROPs) genes and dense granules (GRA) genes, were retrieved from TOXODB, by applying multiple queries by using the tool: (Identify Genes based on Text) (http://www.toxodb.org/toxo/showQuestion.do?questionFullName=GeneQuestions.GenesByTextSearch).

4.2.2. SNP calling

The mutation analysis was performed by Dr. Rachel Brenchley at the University of Salford, using BCFtools (http://samtools.github.io/bcftools/).

The genes of three biologically important families of proteins (SAG1- related sequences (SRS) genes, Rhoptry (ROPs) genes and dense granules (GRA) genes) were identified within the annotated genome file of TgME49 in gff format (retrieved from TOXODB) (http://www.toxodb.org/common/downloads/Current_Release/TgondiiME49/gff/data/) for further analysis as follows:

- **Identification of SAG1- related sequences (SRS) genes:**

  Command:

  grep 'SAG-related' sorted_ToxoDB-12.0_TgondiiME49.gff | grep 'chr' >toxo_SRS_genes.gff

- **Identification of Rhoptry (ROPs) genes:**

  Command:

  grep 'rhoptry' sorted_ToxoDB-12.0_TgondiiME49.gff >toxo_ROP_genes.gff

  Add in the 1 remaining gene that does not have the word 'rhoptry' in its description:

  grep 'toxofilin' sorted_ToxoDB-12.0_TgondiiME49.gff

- **Identification of Dense granule (GRA) genes:**

  Command:

  grep 'dense' sorted_ToxoDB-12.0_TgondiiME49.gff | grep 'gran'>toxo_GRA_genes.gff

A customised set of commands was used to call SNPs within the genes of SRS, ROPs and GRA groups in TgCkUg8 and 9 mapped reads against the type II reference strain genome sequences of *Toxoplasma gondii* (TgME49) in fasta format (retrieved from TOXODB) (http://toxodb.org/common/downloads/Current_Release/TgondiiME49/fasta/data/) and annotated genome file of TgME49 in gff format (retrieved from TOXODB) (http://www.toxodb.org/common/downloads/Current_Release/TgondiiME49/gff/data/) with 3 and 5 reads depths for each sample (see Appendix 2).

4.2.3. Visualization of the coverage and variations within TgCkUg8 and 9

The sorted and indexed BAM files for both TgCkUg8 and 9 were loaded into Artemis (version 16.0.0) (https://www.sanger.ac.uk/resources/software/artemis/) for visualization of the coverage and variations of detected variable genes of both Ugandan strains (TgCkUg8 and 9) against the type II reference strain genome sequences of *Toxoplasma gondii* (TgME49) in fasta format (retrieved from TOXODB) and annotated genome file of TgME49 in gff format (retrieved from TOXODB).

4.3. RESULTS

The genes of three biologically important families of proteins located on the parasite surface (SRS), in the rhoptries (ROPs) and in dense granules (GRA) genes, are not clustered together but distributed across the 14 chromosomes of *T. gondii* (Table 4.4).

Table 4.4: Distribution and number of SAG1- related sequence (SRS) genes, rhoptry (ROPs) genes and dense granules (GRA) genes among 14 chromosomes of *Toxoplasma gondii* ME49

Chromosomes	No. of ROP genes	No. of SRS genes	No. of GRA genes
Ia	4	1	0
Ib	2	4	0
II	1	2	2
III	2	1	2
IV	2	14	2
V	0	32	1
VI	7	28	1
VIIa	5	3	1
VIIb	9	6	0
VIII	3	9	2
IX	4	12	1
X	6	44	4
XI	11	21	1
XII	5	5	1
None	1	2	0
Total	**62**	**111**	**18**

4.3.1. SNPs calling within three biologically important families of proteins

The novel generated sequences (reads) for two Ugandan strains TgCkUg8 and TgCkUg9 (see Chapter 3) were mapped to 14 chromosomes of the type II reference genome TgME49 (http://toxodb.org/common/downloads/Current_Release/TgondiiME49/fasta/data/) and the annotated genome file of TgMg49 in gff format (http://www.toxodb.org/common/downloads/Current_Release/TgondiiME49/gff/data/) to identify SNPs within SAG1- related sequences (SRS) genes and Rhoptry (ROPs) genes and dense granules (GRA) genes across 14 chromosomes of *T. gondii* ME49 at minimum 3 and 5 reads depths.

In TgCkUg8 strain, the coverage percentages for SRS, ROPs and GRA genes are mostly similar at 3 particular reads depths (12.9%, 12.3% and 13.1% respectively) and 5 reads depth (2.2%, 2.1% and 1.9% respectively). In addition, the coverage percentages for SRS, ROPs and GRA genes in TgCkUg9 are also mostly similar at 3 reads depth (49.2%, 49.5% and

64.0% respectively) and 5 reads depth (21.8%, 24.1% and 35.1% respectively) (Table 4.5). This indicates that there is no obvious coverage bias in all three genes families for both strains.

Table 4.5: The total coverage for three families' genes in TgCkUg8 and 9 at 3x and 5x depths against the reference genome TgME49

Genes family	Total length of genes	TgCkUg8				TgCkUg9			
		At 3x depth		At 5x depth		At 3x depth		At 5x depth	
		No. bases covered	Percent coverage (%)	No. bases covered	Percent coverage (%)	No. bases covered	Percent coverage (%)	No. bases covered	Percent coverage (%)
SRS	220875	28585	12.9	4958	2.2	108562	49.2	48190	21.8
ROPs	296091	36387	12.3	6244	2.1	146683	49.5	71100	24.1
GRA	58199	7644	13.1	1100	1.9	37249	64.0	20406	35.1

4.3.1.1.Variations within Surface gene family (SAG1- related sequences (SRS))

Many surface antigens were identified in *Toxoplasma gondii*, which are members of a family of proteins defined by the first member to be described, SAG1, named SAG1- related sequences (SRS) genes. 111 SRS genes were retrieved from TOXODB (see Section 4.3). In order to detect SNPs within 111 genes of the SRS family of both Ugandan strains, the sequence reads of TgCkUg8 and 9 were mapped against the genome sequences and the annotated genome file of the type II reference strains (TgME49) (retrieved from TOXODB) (Table 4.6).

Table 4.6: Variations within SAG1- related sequences (SRS) family genes across 14 chromosomes of TgME49 for two Ugandan strains (TgCkUg8 and 9)

Chromosomes	TgCkUg8						TgCkUg9					
	At 3x depth			At 5x depth			At 3x depth			At 5x depth		
	No. bases covered	SNPs		No. bases covered	SNPs		No. bases covered	SNPs		No. bases covered	SNPs	
		No.	SNPs /kb		No.	SNPs /kb		No.	SNPs /kb		No.	SNPs /kb
Ia	82	0	0	0	0	0	729	0	0	98	0	0
Ib	1030	0	0	26	0	0	6663	1	0.2	2419	0	0
II	1186	0	0	14	0	0	1756	0	0	765	0	0
III	176	0	0	0	0	0	788	0	0	157	0	0
IV	2244	21	9.4	481	5	10.4	9167	3	0.3	1798	3	1.7
V	909	0	0	0	0	0	2215	0	0	489	0	0
VI	802	0	0	28	0	0	7305	2	0.3	3393	1	0.3
VIIa	1235	0	0	127	0	0	6389	2	0.3	3175	1	0.3
VIIb	1784	4	2.2	813	0	0	3162	7	2.2	978	5	5.1
VIII	2474	0	0	331	0	0	6847	1	0.1	2588	1	0.4
IX	2067	1	0.5	283	0	0	17323	3	0.2	9666	2	0.2
X	6000	3	0.5	1174	0	0	29282	39	1.3	17865	22	1.2
XI	5205	3	0.6	898	1	1.1	11855	1	0.1	3317	0	0
XII	3391	0	0	783	0	0	5081	1	0.2	1482	0	0
Total		**32**			**6**			**60**			**35**	

Many SNPs were identified within SRS genes of TgCkUg8 and 9 strains. In TgCkUg8, the total number of SNPs that were identified within some members of 111 genes of the SRS family were 32 within 9 genes at a depth of minimum 3 reads (Table 4.7).

Table 4.7: Coverage, detected number of SNPs and estimated SNPs/kb in SAG1- related sequences (SRS) of TgCkUg8 at 3x and 5x depth compared with the reference genome TgME49

			TgCkUg8							
Chromosomes	Genes	Length	At 3x depth				At 5x depth			
			No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb	No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb
IV	SRS15A	1179	437	37.07	19	43.5	272	23.07	3	11.1
	SRS16B	1889	255	13.50	2	7.8	99	5.24	2	20.2
VIIb	SRS26A	1332	339	25.45	4	11.8	-	-	-	-
IX	SRS38B	1652	437	26.45	1	2.3	-	-	-	-
X	SRS40B	1575	148	9.40	1	6.8	-	-	-	-
	SRS48K	1854	1223	65.97	2	1.6	-	-	-	-
XI	SRS51	2771	316	11.40	1	3.2	-	-	-	-
	SRS55F	2323	578	24.88	1	1.7	215	9.26	1	4.7
	SRS52C	4009	344	8.58	1	2.9	-	-	-	-
Total	**9 genes**	**18584**	**4077**	**21.9**	**32**	**7.8**	**586**	**10.9**	**6**	**10.2**

In TgCkUg9, the total number of SNPs that were identified within some members of 111 genes of the SRS family was 60 within 18 genes at a depth of a minimum of 3 reads (Table 4.8).

Table 4.8: Coverage percentages, detected number of SNPs and estimated SNPs/kb in SAG1- related sequences (SRS) of TgCkUg9 at 3x and 5x depths compared with the reference genome TgME49

TgCkUg9										
Chromosomes	Genes	Length	At 3x depth				At 5x depth			
			No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb	No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb
Ib	SRS12B	2873	2052	71.42	1	0.5	-	-	-	-
IV	SRS17B	5401	1446	26.77	1	0.7	332	6.15	1	3.1
IV	SRS19C	1430	384	26.85	1	2.6	118	8.25	1	8.5
IV	SRS19F	1728	983	56.89	1	1.1	470	27.20	1	2.1
VI	SRS22I	740	426	57.57	1	2.3	219	29.59	1	4.6
VI	SRS23	1877	526	28.02	1	1.9	-	-	-	-
VIIa	SRS39	2296	2105	91.68	2	0.9	1718	74.83	1	0.6
VIIb	SRS26A	1332	952	71.47	6	6.3	684	51.35	4	5.8
VIIb	SRS26C	1455	833	57.25	1	1.2	112	7.70	1	8.9
VIII	SRS29C	2752	1831	66.53	1	0.5	679	24.67	1	1.5
IX	SRS38A	1934	1934	100.00	1	0.5	1897	98.09	1	0.5
IX	SRS44	8802	5192	58.99	2	0.4	1736	19.72	1	0.6
X	SRS46	3206	2279	71.09	24	10.5	1394	43.48	13	9.3
X	SRS47B	2102	1758	83.63	2	1.1	898	42.72	1	1.1
X	SRS48Q	1907	1017	53.33	10	9.8	587	30.78	5	8.5
X	SRS48K	1854	1601	86.35	3	1.9	1376	74.22	3	2.2
XI	SRS52C	4009	1506	37.57	1	0.7	-	-	-	-
XII	SRS57 (**SAG3**)	3785	2924	77.25	1	0.3	-	-	-	-
Total	**18 genes**	**49483**	**29749**	**60.1**	**60**	**2.1**	**11081**	**30.0**	**35**	**3.2**

To increase confidence in the SNPs, they were called against the reference genome sequence at a minimum of 5 reads depth for all gene candidates within the SRS family of both Ugandan strains. The total number of SNPs that were identified within this group of genes decreased to 6 SNPs within 3 genes in TgCkUg8 (Table 4.7), and 35 within 14 genes in the TgCkUg9 strain (Table 4.8). In TgCkUg8, the SRS15A gene has the highest estimated SNPs/kb rate by 43.5 SNPs/kb at 3 reads depth, whereas SRS16B has the highest rate by 20.2 SNPs/kb at 5x depth. In TgCkUg9, SRS46 on chromosome X was the most variable gene with an estimated SNPs/kb rate of 10.5 and 9.3 at 3x and 5x depths respectively. There

are no similar variable SRS genes detected in both Ugandan strains, which indicates that they are not identical.

4.3.1.2.Variations within rhoptry gene family (ROPs)

Rhoptries are groups of proteins that are stored in rhoptry organelles of *T. gondii* and released during the invasion in the host cell to facilitate intracellular interactions with it. 62 ROP genes were retrieved from TOXODB (see Section 4.3). After mapping the sequences of TgCkUg8 and 9 strains against the annotated sequences of the reference strain genome (TgME49) (retrieved from TOXODB), variations were detected within some members of 62 genes of the ROPs family (Table 4.9).

Table 4.9: Variations within rhoptry (ROPs) family genes across 14 chromosomes of TgME49 for two Ugandan strains (TgCkUg8 and 9)

Chromosomes	TgCkUg8						TgCkUg9					
	At 3x depth			At 5x depth			At 3x depth			At 5x depth		
	No. bases covered	SNPs		No. bases covered	SNPs		No. bases covered	SNPs		No. bases covered	SNPs	
		No.	SNPs /kb		No.	SNPs /kb		No.	SNPs /kb		No.	SNPs /kb
Ia	1835	4	2.2	679	1	1.5	5934	20	3.4	2979	19	6.4
Ib	38	0	0	0	0	0	1269	0	0	10	0	0
II	1250	0	0	126	0	0	7417	0	0	2928	0	0
III	336	0	0	0	0	0	1111	1	0.9	383	1	2.6
IV	318	0	0	22	0	0	2502	0	0	1338	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0
VI	3781	50	13.2	1426	16	11.2	11757	26	2.2	5529	21	3.8
VIIa	1770	0	0	73	0	0	23431	1	0.1	20890	1	0.1
VIIb	1921	1	0.5	117	0	0	13432	2	0.1	5047	0	0
VIII	2856	15	5.3	466	15	32.2	7057	15	2.1	2137	15	7.1
IX	3404	0	0	417	0	0	12553	1	0.1	3895	1	0.3
X	4317	17	3.9	790	6	7.6	12753	27	2.1	8550	26	3.1
XI	10926	1	0.1	901	0	0	37261	6	0.2	12528	1	0.1
XII	3636	17	4.7	1227	6	4.9	10206	32	3.1	4886	31	6.3
Total		**105**			**44**			**131**			**116**	

Many SNPs were identified within the ROP genes of TgCkUg8 and 9 strains. Overall, genes of the ROP family in these Ugandan strains were more variable compared with SRS genes. In TgCkUg8, the total number of SNPs identified within some members of 62 genes of the ROPs family was 105 within 10 genes at a depth of a minimum of 3 reads (Table 4.10).

Table 4.10: Coverage percentages, detected number of SNPs and estimated SNPs/kb in rhoptry (ROPs) genes of TgCkUg8 at 3x and 5x depths compared with the reference genome TgME49

TgCkUg8										
Chromosomes	Genes	Length	At 3x depth				At 5x depth			
			No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb	No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb
Ia	ROP7	2547	1161	45.58	4	3.4	679	26.66	1	1.5
VI	ROP19A	2810	1748	62.21	22	12.6	911	32.42	15	16.5
VI	ROP19B	1656	914	55.19	28	30.6	271	16.36	1	3.7
VIIb	RON10	4760	304	6.39	1	3.3	-	-	-	-
VIII	RON4	8660	1036	11.96	15	14.5	373	4.31	15	40.2
X	ROP8	3052	1599	52.39	15	9.4	678	22.21	6	8.8
X	ROP2A	1729	716	41.41	2	2.8	-	-	-	-
XI	RON9	10485	1050	10.01	1	0.9	-	-	-	-
XII	**ROP5**	1650	1434	86.91	4	2.8	886	53.70	3	3.4
XII	ROP	4105	990	24.12	13	13.1	182	4.43	3	16.5
Total	**10 genes**	**41454**	**10952**	**26.4**	**105**	**9.6**	**3980**	**16.3**	**44**	**11.1**

In TgCkUg9, the total number of SNPs that were identified within some members of the 62 genes of the ROPs family was 131 within 21 genes at a depth of a minimum of 3 reads (Table 4.11).

Table 4.11: Coverage percentages, detected number of SNPs and estimated SNPs/kb in rhoptry (ROPs) genes of TgCkUg9 at 3x and 5x depths compared with the reference genome TgME49

TgCkUg9										
Chromosomes	Genes	Length	At 3x depth				At 5x depth			
			No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb	No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb
Ia	ROP7	2547	2367	92.93	19	8.1	1993	78.25	19	9.5
Ia	ROP4	1308	802	61.31	1	1.2	-	-	-	-
III	ROP24	3040	1111	36.55	1	0.9	383	12.60	1	2.6
VI	ROP38	2572	1863	72.43	1	0.5	845	32.85	1	1.2
VI	ROP19A	2810	2498	88.90	20	8.1	1903	67.72	18	9.5
VI	ROP19B	1656	1577	95.23	5	3.2	1448	87.44	2	1.4
VIIa	ROP25	4808	4808	100.00	1	0.2	4655	96.82	1	0.2
VIIb	ROP20	1941	693	35.70	2	2.9	-	-	-	-
VIII	RON4	8660	3116	35.98	15	4.8	1452	16.77	15	10.3
IX	RON8	16369	5383	32.89	1	0.2	1500	9.16	1	0.7
X	ROP11	3806	3383	88.89	1	0.3	2653	69.71	1	0.4
X	RON3	11893	3328	27.98	1	0.3	-	-	-	-
X	ROP8	3052	2613	85.62	16	6.1	2521	82.60	16	6.3
X	ROP2A	1729	1696	98.09	9	5.3	1622	93.81	9	5.5
XI	ROP1	4246	589	13.87	1	1.7	-	-	-	-
XI	RON5	21427	8371	39.07	1	0.1	-	-	-	-
XI	ROP27	9570	8346	87.21	2	0.2	-	-	-	-
XI	ROP14	9838	3059	31.09	1	0.3	-	-	-	-
XI	ROP	7479	4819	64.43	1	0.2	2357	31.51	1	0.4
XII	**ROP5**	1650	1648	99.88	9	5.5	1469	89.03	8	5.4
XII	ROP	4105	2671	65.07	23	8.6	1971	48.01	23	11.7
Total	**21 genes**	**124506**	**64741**	**51.9**	**131**	**2.1**	**26772**	**41.6**	**116**	**4.3**

To increase confidence in the SNPs, they were called against the reference genome sequence at a minimum of 5 reads depth for all gene candidates within the ROPs families of both Ugandan strains. The total number of SNPs that were identified within this group of genes decreased to 44 SNPs within 7 genes in the TgCkUg8 (Table 4.10), and 116 within 14 genes in the TgCkUg9 strain (Table 4.11). All variable 7 ROPs genes identified in the TgCkUg8 strain were also identified as variable genes in the TgCkUg9 strain.

4.3.1.3.Variations within dense granules gene family (GRA)

Dense granules are group of proteins in *T. gondii* that are exocytosed during or after invasion of the host cell. 18 GRA genes were retrieved from TOXODB (see Section 4.3). Variations were detected within some members of 18 genes of the GRA family through alignment of sequences of TgCkUg8 and 9 strains against the annotated sequences of the reference strain

genome (TgME49) (retrieved from TOXODB) (Table 4.12).

Table 4.12: Variations within dense granules (GRA) family genes across 14 chromosomes of TgME49 for two Ugandan strains (TgCkUg8 and 9)

Chromosomes	TgCkUg8						TgCkUg9					
	At 3x depth			At 5x depth			At 3x depth			At 5x depth		
	No. bases covered	SNPs		No. bases covered	SNPs		No. bases covered	SNPs		No. bases covered	SNPs	
		No.	SNPs /kb		No.	SNPs /kb		No.	SNPs /kb		No.	SNPs /kb
Ia	0	0	0	0	0	0	0	0	0	0	0	0
Ib	0	0	0	0	0	0	0	0	0	0	0	0
II	63	0	0	26	0	0	2080	2	0.9	388	0	0
III	257	0	0	0	0	0	3711	1	0.3	2065	1	0.5
IV	1344	1	0.7	90	0	0	3006	0	0	1538	0	0
V	38	0	0	0	0	0	1355	0	0	681	0	0
VI	76	0	0	0	0	0	1459	0	0	292	0	0
VIIa	710	0	0	92	0	0	4614	0	0	3651	0	0
VIIb	0	0	0	0	0	0	0	0	0	0	0	0
VIII	402	0	0	0	0	0	3809	0	0	739	0	0
IX	280	0	0	0	0	0	4346	1	0.2	4325	1	0.2
X	3694	10	2.7	892	9	10.1	8282	9	1.1	3025	9	2.9
XI	256	0	0	0	0	0	982	0	0	200	0	0
XII	524	0	0	0	0	0	3605	0	0	3502	0	0
Total		**11**			**9**			**13**			**11**	

Several SNPs were identified within GRA genes of the TgCkUg8 and 9 strains. GRA proteins had the lowest number of SNPs identified in some genes of this family, compared with SRS and ROPs genes. In TgCkUg8, the total number of SNPs identified within some members of 18 genes of the GRA family was 11 within 3 genes at a depth of a minimum of 3 reads (Table 4.13).

Table 4.13: Coverage percentages, detected number of SNPs and estimated SNPs/kb in dense granule (GRA) of TgCkUg8 at 3x and 5x depths compared with the reference genome TgME49

TgCkUg8										
Chromosomes	Genes	Length	At 3x depth				At 5x depth			
			No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb	No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb
IV	GRA11	2687	644	23.97	1	1.6	-	-	-	-
X	GRA3	5056	1079	21.34	9	8.3	385	7.61	9	23.4
X	GRA15	3461	866	25.02	1	1.2	-	-	-	-
Total	**3 genes**	**11204**	**2589**	**23.1**	**11**	**4.2**	**385**	**7.61**	**9**	**23.4**

In TgCkUg9, the total number of SNPs identified within some members of 18 genes of the GRA family was similar to the previous Ugandan strain, with 13 within 4 genes at a depth of a minimum of 3 reads (Table 4.14).

Table 4.14: Coverage percentages, detected number of SNPs and estimated SNPs/kb in dense granule (GRA) of TgCkUg9 at 3x and 5x depths compared with the reference genome TgME49

TgCkUg9										
Chromosomes	Genes	Length	At 3x depth				At 5x depth			
			No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb	No. bases covered	Percent coverage (%)	No. of SNPs	SNPs /kb
II	GRA32	3873	1250	32.27	2	1.6	-	-	-	-
III	GRA12	3994	3475	87.01	1	0.3	2053	51.40	1	0.5
IX	GRA12	4346	4346	100.00	1	0.2	4325	99.52	1	0.2
X	GRA3	5056	2887	57.10	9	3.1	711	14.06	9	12.7
Total	**4 genes**	**17269**	**11958**	**69.2**	**13**	**1.1**	**7089**	**52.9**	**11**	**1.6**

To increase confidence in the SNPs, they were called against the reference genome sequence at a minimum of 5 reads depth for all gene candidates within the GRA family of both Ugandan strains. The total number of SNPs identified within this group of genes decreased to 9 SNPs within one gene in TgCkUg8 (Table 4.13), and 11 within 3 genes in the TgCkUg9 strain (Table 4.14). The GRA3 gene in chromosome X has the most variability within this group of genes, with an estimated SNPs/kb rate of 23.4 and 12.7 at 5x depth in TgCkUg8

and 9 respectively. In addition, GRA3 is the only variable gene within the GRA family that was detected in both Ugandan strains.

4.3.2. The variable genes within three biologically important families of proteins detected in both TgCkUg8 and 9 strains

Review of the variable genes detected in both Ugandan strains compared to the reference strain TgME49 indicates that they are not identical to the reference strain. In SRS genes, only 9 variable genes were detected in TgCkUg8 at 3x depth which decreased to 3 genes at 5x depth out of 111 SRS genes, while in TgCkUg9 there are only 18 candidates at 3x depth decreasing to 14 genes at 5x depth out of a total of 111 SRS genes. However, no similar variable SRS genes were identified in both Ugandan strains, which indicates that they are not identical. In addition, only 10 variable ROPs genes were identified at 3x depth which declined to 7 candidates at 5x depth out of a total 62 ROPs genes in the TgCkUg8 strain, whereas only 21 genes at 3x depth, reducing to 14 candidates at 5x depth, were identified in the TgCkUg9 strain. In the GRA genes family, 3 variable genes detected at 3x depth declined to one gene at 5x depth out of a total of 18 GRA genes in TgCkUg8, while 4 genes at 3x depth decreased to 3 candidates at 5x depth in the TgCkUg9 strain. However, most of these variable genes were not identified as similar in both Ugandan strains, which indicates that they are not identical.

By choosing the 5x coverage data as indicative, to increase confidence of SNPs within variable genes, no similar variable SRS genes were identified in both Ugandan strains. In contrast, particular 7 of the recognized variable ROPs genes (ROP7, ROP19A, ROP19B, ROP8, ROP5, RON4 and ROP on chr. XII) and one GRA gene (GRA3) were detected in both TgCkUg8 and 9 strains (Table 4.15). These particular 8 variable genes, which are the only variable genes that were detected in both Ugandan strains at 5x depth, were chosen to evaluate their ability to discriminate the local type II strains through detection of local variation within the related strains (see following Section).

Table 4.15: The variable genes within three biologically important families of proteins (SAG1-related sequences (SRS) genes, rhoptry (ROPs) genes and dense granules (GRA) genes) which are detected in both Ugandan strains TgCkUg8 & 9

Genes	Chromosomes	No. of SNPs in TgCkUg8 At 5x depth	No. of SNPs in TgCkUg9 At 5x depth
ROP7	Ia	1	19
ROP19A	VI	15	18
ROP19B	VI	1	2
ROP8	X	6	16
ROP5	XII	3	8
ROP	XII	3	23
RON4	VIII	15	15
GRA3	X	9	9

To visualise coverage and variations within these 8 genes of both Ugandan strains, Artemis 16.0.0 (https://www.sanger.ac.uk/resources/software/artemis/) was used. The sorted and indexed BAM files for both TgCkUg8 and 9 were loaded into Artemis for visualization of the coverage and variations of these 8 genes of both Ugandan strains against the genome sequences and the annotated genome file of the type II reference strain TgME49 (retrieved from TOXODB).

In Figure 4.1, it is clear from the coverage graphs of TgCkUg8 & 9 (mapped reads) to the reference genome (TgME49) (in red and blue respectively) that the sequence reads (represented by short horizontal lines) of both strains covered most of the ROP5 gene, and the coverage is greater in TgCkUg9 compared to the TgCkUg8 strain. This confirms the estimated coverage percentages in TgCkUg8 at 86.91% and 53.70% at 3 and 5 depths respectively (Table 4.11), and TgCkUg9 at 99.88% and 89.03% at 3 and 5 depths respectively (Table 4.12). Polymorphisms in ROP5 gene will lead to codon changes which may in turn lead to amino acids modifications when non synonymous mutation occurs. SNPs against the reference genome (TgME49) are represented by vertical red marks showing the reads depths of these polymorphisms.

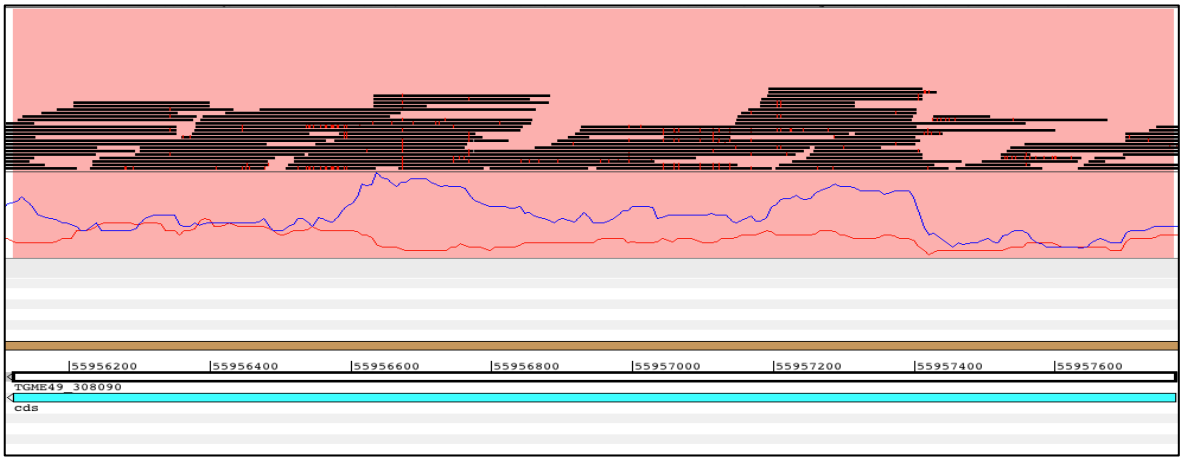

Figure 4.1: The coverage and variations within ROP5 gene in both TgCkUg8 and 9 strains (by use of Artemis).
The coverage graphs of TgCkUg8 & 9 (mapped reads) to the reference genome (TgME49) (in red and blue respectively). Horizontal lines represent the mapped reads of both TgCkUg8 and 9 strains, and vertical red marks represent SNPs within these sequences against the reference genome (TgME49)

From Figure 4.2 below, it is evident that sequence reads of both Ugandan strains were mapped properly to five of these variable genes (ROP7, ROP19A, ROP19B, ROP8 and ROP in chr. XII), which supports the reliability and confidence of SNPs within these genes. It is clear that sequence reads of TgCkUg8 and 9 were mapped repetitively in RON4 and GRA3 genes. It can be seen that there is an accumulation of repetitive reads in one region within RON4 and GRA3 genes that includes most SNPs, which suggests insufficient reliability of polymorphisms detected in these two genes. Therefore, RON4 and GRA3 genes were excluded from the list of variable genes within the three biologically important families of proteins.

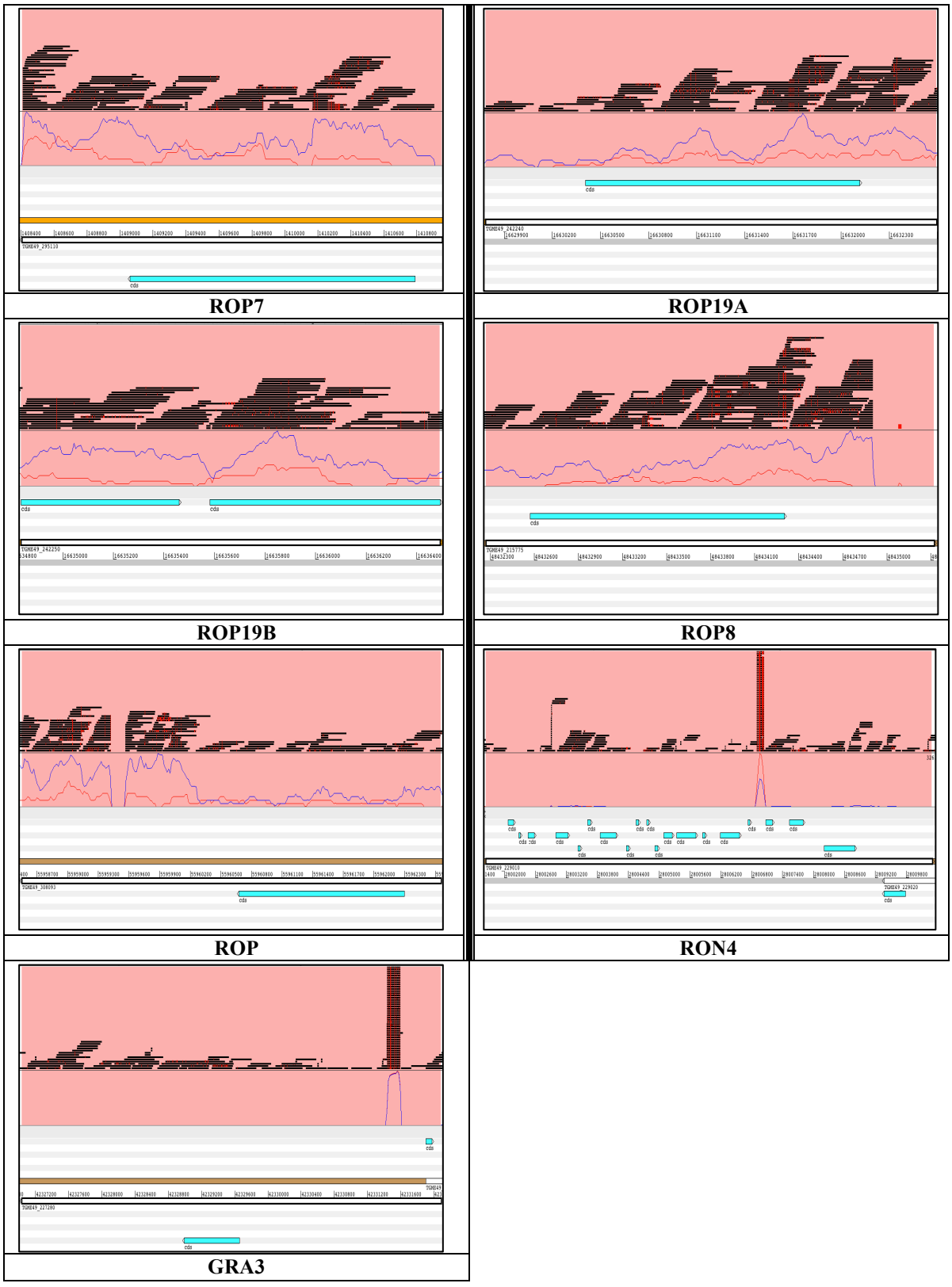

Figure 4.2: The coverage and variations within seven variable genes of three biologically important families of proteins that were detected in both TgCkUg8 and 9 strains (by use of Artemis).
The coverage graphs of TgCkUg8 & 9 (mapped reads) to the reference genome (TgME49) (in red and blue respectively). Horizontal lines represent the mapped reads of both TgCkUg8 and 9 strains, and vertical red marks represent SNPs within these sequences against the reference genome (TgME49).

To check the reliability of SNPs within all variable genes in the three families (except the affected two genes RON4 & GRA3), coverage and SNP plots were generated for all variable genes by use of Artemis (version16.0.0). It was shown that patterns of coverage in these variable genes are normal which confirm the reliability of polymorphisms of these candidates.

By focusing more on the distribution of SNPs within the ROP5 gene located on chromosome XII (as an example) in TgCkUg8 and 9 strains at 5 reads depth compared with type II (ME49) and III (VEG) reference strains of *T. gondii* (Figure 4.3), it is clear that 3 SNPs in the TgCkUg8 strain are novel SNPs (different from both references) and 22 SNPs have a type II background (identical to type II reference - ME49). In TgCkUg9, 5 SNPs are novel, while 3 SNPs have a type III background (identical to type III reference – VEG), and 17 SNPs have a type II background. Detection of SNPs in the ROP5 gene in both Ugandan strains with 3 SNPs in TgCkUg8 and 8 polymorphisms in the TgCkUg9 strain evident that this virulence associated gene is highly polymorphic candidate.

Strains	566861	566863	566888	566893	566903	566908	566909	566954	566956	567143	567158	567192	567195	567269	567274	567368	567369	567596	567645	567739	567806	567811	568020	568180	568188
ME49	T	G	C	A	T	A	G	T	G	T	T	T	A	A	C	A	G	C	G	A	C	C	G	C	G
TgCkUg8	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	A	*	T	T	*	*	*
TgCkUg9	*	*	*	*	*	*	*	*	*	*	*	A	C	*	T	G	*	A	*	*	T	T	C	*	*
VEG	C	A	G	G	C	G	C	A	C	C	G	A	C	T	*	G	C	*	*	G	*	*	*	G	A

Figure 4.3: The distribution of SNPs within the ROP5 gene in TgCkUg8 and 9 strains compared with type II and III reference strains (ME49 and VEG respectively).
Green SNPs represent a type II background (identical to type II reference - ME49), blue SNPs represent a type III background (identical to type III reference – VEG), and red SNPs represent novel SNPs (different from both references). Similarities to type II reference sequence (TgME49) are indicated by (*). Numerical positions refer to sites in the published sequences of the (TgME49) reference strain (http://toxodb.org/common/downloads/Current_Release/TgondiiME49/fasta/data/)

The local variations in TgCkUg8 and 9 strains within ROP5 genes lead to coding changes and modification of the protein structure by a change in amino acid composition (Figure 4.4). In the TgCkUg8 strain, 2 novel SNPs resulted in amino acid changes from aspartate (D) and Threonine (T), to Asparagine (N) and Methionine (M) respectively in comparison with the TgME49 reference strain. 3 novel SNPs in the TgCkUg9 strain led to amino acid modifications from Threonine (T) (two amino acids) and Glycine (G) to Methionine (M) (two amino acids) and Arginine (R) respectively in comparison with the TgME49 reference strain. Both Ugandan strains have the same changes in amino acid (in locus amino acid position of 364 in ROP5 gene in TgME49 sequence) from Threonine (T) to Methionine (M) in comparison with the TgME49 reference strain (Figure 4.4). One polymorphism that has a type III background in TgCkUg9 strain resulted in a change of amino acid from Phenylalanine (F) to Isoleucine (I) compared with the TgME49 reference strain, which is similar to the type III reference strain (VEG).

Strains	Locus amino acid positions in ROP5 gene in TgME49 sequence													
	48	56	58	61	146	158	185	217	309	340	364	434	487	490
ME49	S	H	Y	V	H	F	T	E	D	H	T	G	A	E
TgCkUg8	*	*	*	*	*	*	*	*	N	*	M	*	*	*
TgCkUg9	*	*	*	*	*	I	M	*	*	*	M	R	*	*
VEG	N	Q	C	A	Q	I	*	Q	*	R	*	*	G	K

Figure 4.4: Local allelic variations in TgCkUg8 and 9 strains leading to amino acid changes in a rhoptry protein – ROP5 located on chromosome XII, compared with type II and III reference strains (ME49 and VEG respectively)
Similarities to type II reference sequence (ME49) are indicated by (*)
(Using (MEGA) version 5.1 (http://www.megasoftware.net))

4.3.3. Variations among type II strains of *Toxoplasma gondii*

In Chapter 2, it was reported that the current 10 RFLP markers (SAG1, SAG2, GRA6, BTUB, L358, C22-8, C29-2, PK1, and Apico (except SAG3, not used)) were insufficient to clearly discriminate between all type 2 and 12 strains of *T. gondii* (Figure 2.9, Tables 2.10 and 4.16).

Table 4.16: Number of SNPs among type II and type 12 strains in addition to type I (GT1) and III (VEG) reference genomes of *Toxoplasma gondii* against the reference genome TgME49 using 9 RFLP genetic markers.
These data were retrieved from TOXODB by applying multiple queries for each marker by using the tools: (Identify SNPs based on Gene IDs) or (Identify SNPs based on Genomic Location-chromosome.
(http://www.toxodb.org/toxo/showQuestion.do?questionFullName=SnpQuestions.NgsSnpsByGeneIds)
(http://www.toxodb.org/toxo/showQuestion.do?questionFullName=SnpQuestions.NgsSnpsByLocation)

Genetic markers		**No. of SNPs**						
Genes	**Gene ID**	**Type I (Reference)**	**Type III (Reference)**	**Type II strains**		**Type 12 strains**		
		ME49-GT1	**ME49-VEG**	**ME49 - B73**	**ME49 - PRU**	**ME49 - ARI**	**ME49 - B41**	**ME49 -RAY**
SAG1	TGME49_233460	10	0	0	0	0	0	0
SAG2	TGME49_271050	15	16	0	0	0	0	0
SAG3	TGME49_308020	77	80	0	1	1	0	0
GRA6	TGME49_275440	32	40	0	0	47	0	47
BTUB	TGME49_266960	45	45	0	0	18	0	8
PK1	TGME49_243500	36	48	48	0	2	0	14
C22-8	TGME49_chrIb 1854961-1855560	2	2	0	0	0	0	0
C29-2	TGME49_252890	39	38	0	0	3	0	1
L358	TGME49_285780	132	135	135	0	137	115	109

However, globally, it was shown that variations between different *T. gondii* strains from several regions in the world could be detected by using several genetic markers, including these 10 RFLP markers (Su et al., 2010, Su et al., 2012). In the current Chapter, analysis was performed on three biologically relevant genes families, and identified many variable genes within these groups. Six variable genes were identified in both Ugandan strains as reported in Section 4.3.2**.** These six variable genes have the ability to discriminate and increase resolution amongst the local type II and 12 strains through detection of local variations within these related strains (Table 4.17 and Figure 4.5). For instance, it is clear that using these loci is adequate to differentiate two type II strains (TgCkUg8 and PRU) from the reference strain (TgME49) that were identical to this reference strain via using RFLP markers (Figure 2.9**).** It is interesting to look more widely at these polymorphic markers and add them into the MLST analysis.

Table 4.17: Number of SNPs among type II and type 12 strains in addition to type I (GT1) and III (VEG) reference genomes of *Toxoplasma gondii* against the type II reference genome (TgME49) using six genetic markers (which are detected in this study as the variable genes within both Ugandan strains).
Data were obtained from this study (for TgCkUg8 & 9 strains) and retrieved from TOXODB (for the other strains) by applying multiple queries for each marker by using the tools: (Identify SNPs based on Gene IDs).
(http://www.toxodb.org/toxo/showQuestion.do?questionFullName=SnpQuestions.NgsSnpsByGeneIds)

Variable genes	No. of SNPs								
Gene Gene ID	Type I (Reference)	Type III (Reference)	Type II strains				Type 12 strains		
	GT1	VEG	TgCkUg8	TgCkUg9	B73	PRU	ARI	B41	RAY
ROP7 TGME49_295110	0	0	1	19	0	0	25	0	4
ROP19A TGME49_242240	8	2	15	18	2	2	14	1	6
ROP19B TGME49_242250	3	0	1	2	0	0	11	0	0
ROP8 TGME49_215775	29	31	6	16	0	4	48	0	40
ROP5 TGME49_308090	23	19	3	8	0	0	0	0	1
ROP TGME49_308093	69	70	3	23	1	3	4	1	4

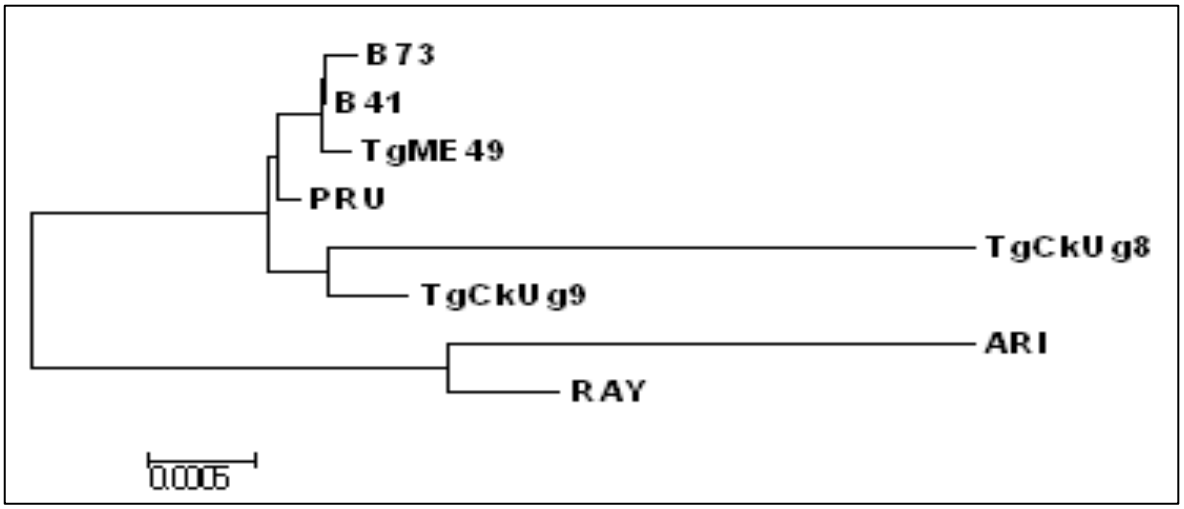

Figure 4.5: Construction of Neighbor-Joining tree using concatenated DNA sequences of ROP7, ROP19A, ROP19B, ROP8, ROP5 and ROP (on chr. XII) loci from both Ugandan strains (TgCkUg8 and 9), the type II reference strain (TgME49), type II strains (PRU and B73) and type 12 strains (B41, ARI and RAY) of *Toxoplasma gondii*
(By use of MEGA 5.1 (http://www.megasoftware.net))

4.4. DISCUSSION

In *Toxoplasma gondii*, there are key biologically important sets of genes, which are involved in mediating intracellular lifestyle of this parasite. Three of these gene groups include surface, rhoptry and dense granule gene families. Some members of these three families were seen to be associated with virulence in mice, including SAG1, SAG3, ROP16, ROP18 and ROP5 (Howe et al., 1996; Saeij et al., 2006; Taylor et al., 2006).

In this study, polymorphisms were identified within some members of these coding gene families in both Ugandan strains (TgCkUg8 and 9) compared with the reference strain (TgME49), indicating that TgCkUg8 and 9 strains are neither identical to each other nor to the reference strain (TgME49).

It is useful to look at SNPs detected at a minimum depth of 5 reads. In a study conducted by Bontell et al. (2009) for whole genome sequencing of a Ugandan strain (TgCkUg2) by using the 454 platform sequencing method, the validity of SNPs was confirmed by a minimum of 3 reads depth within this genome that had a total coverage of 84.09% of the reference genome TgME49 (61.6 Mbp) at 4x depth. As described in Chapter 3, the total coverage percentage was lower in TgCkUg8 sequences, with only 12.0% of the type II reference strain genome of *T. gondii* - TgME49 (62.5Mbp) at 3x depth, and was relatively lower in TgCkUg9 with 57.15% and 30.4% at 3x depth. In addition, some SNPs that were detected in members of the three genes families at a minimum of 3 reads depth disappeared when SNPs were called with a minimum of 5x depth in both Ugandan strains (Tables 4.9, 10, 11, 12, 13 and 14**).** These observations suggest that the level of confidence is low in SNPs called at a minimum of 3 reads depth. Thus, in TgCkUg8 and 9, SNPs with a minimum depth of 5 reads were assumed to be accurate with a high level of confidence.

There is no bias in the overall coverage of gene families at both 3 and 5 reads depths in both Ugandan strains, as indicated from mostly similar coverage percentages figures for each strain (Table 4.8). At 5x depth, the coverage percentages for SRS, ROPs and GRA genes in TgCkUg8 are 2.2%, 2.1% and 1.9% respectively, while in TgCkUg9 they are 49.2%, 49.5% and 64.0% respectively. As explained earlier, SNPs in TgCkUg8 and 9 which were considered valid by minimum depth of 5 reads can be assumed to be accurate with a high level of confidence. The ROP genes appear to be the most variable candidates among three families, with a total number of SNPs of 44 within 7 rhoptries in TgCkUg8 (Table 4.11), and

116 within 14 ROPs of the TgCkUg9 strain (Table 4.12). The variable genes in the SRS family were confirmed, with 6 SNPs within 3 SRS members in TgCkUg8 (Table 4.9), and 35 polymorphisms in 14 candidates in TgCkUg9 (Table 4.10). In GRA genes, 9 SNPs were detected within only one gene (GRA3) in TgCkUg8 (Table 4.13), while 11 polymorphisms were identified within 3 dense granules in the TgCkUg9 strain (Table 4.13). Bontell et al. (2009) identified fifteen genes with at least two SNPs in the coding sequences of the TgCkUg2 strain that involved two ROPs genes (ROP5 and ROP29) and two SRS genes (SRS16B and SRS26A). In this study, SNPs were detected in the ROP5 gene in both Ugandan strains with 3 SNPs in TgCkUg8 and 8 polymorphisms in the TgCkUg9 strain. In addition, 2 SNPs were identified within the SRS16B gene in TgCkUg8, and 4 polymorphisms within the SRS26A gene in TgCkUg9 strain. Additionally, in order to identify genes with high levels of variation in the TgCkUg2 strain, 46 genes were identified from a high throughput dN/dS screen (the ratio of non-synonymous to synonymous mutations) which include four GRA genes (GRA3, 6, 7 and 8), one ROP gene (ROP7) and one SRS candidate (SRS16B) (Bontell et al., 2009). In our study, SNPs were identified in GRA3 and ROP7 genes in both Ugandan strains with 9 SNPs within GRA3 in both TgCkUg8 and 9 strains, while 1 and 19 polymorphisms were identified within the ROP7 gene in TgCkUg8 and 9 respectively.

It was shown in Chapter 2 that, although the ability of several genetic markers including the current 10 RFLP genetic markers (SAG1, SAG2, GRA6, BTUB, L358, C22-8, C29-2, PK1, and Apico except SAG3 which was not used) to detect global variations within different *T. gondii* strains collected from various regions in the world (Su et al., 2010; Su et al., 2012), these RFLP markers were unable to differentiate between type II and 12 strains of *T. gondii* (Figure 2.9, Tables 2.10 and 4.16).

In the current study, six variable genes were recognized via analysis of three biologically relevant gene families and were used to differentiate the type II and 12 strains of *T. gondii* including both Ugandan strains (Table 4.15). It is clear that these six markers (ROP7, ROP19A, ROP19B, ROP8, ROP5 and ROP (on chr. XII) improved the resolution power to detect variations among local type II strains of *T. gondii*, which confirms that new polymorphisms are useful in the analysis of the genetic structure of the parasite population. It is recommended that further study of these polymorphic markers is carried out and they are added into the MLST analysis of *T. gondii*, especially between local strains.

In a study conducted by Saeij et al. (2006), four virulence-associated loci were identified in an experimental cross between type II and III strains of *T. gondii*, which were located on chromosomes VIIa (ROP18), VIIb (ROP16) and XII (ROP5 and SAG3). SNPs were detected in two virulence-associated genes (ROP5 and SAG3) in TgCkUg8 and 9 strains. Three novel SNPs were identified within the ROP5 gene in TgCkUg8 at 5x depth, while 5 novel polymorphisms were detected in the TgCkUg9 strain at a minimum 5 reads depth (Figure 4.3). In addition, only one SNP was identified within the SAG3 gene in theTgCkUg9 strain at 3 reads depth (Table 4.10). No polymorphisms were detected within ROP18 and 16 in both Ugandan strains, which suggests that both TgCkUg8 and 9 strains are identical to the type II reference strain (TgME49) in these two virulence-associated genes. The local variations in both TgCkUg8 and9 strains within the virulence-associated gene ROP5 resulted in modification of the protein structure by a change in amino acid composition compared with the type II reference strain (TgME49) (Figure 4.4). Two novel polymorphisms in TgCkUg8 and three in TgCkUg9 led to amino acids modifications, while both strains experienced similar changes that were seen to take place in one amino acid (in locus amino acid position of 364 in ROP5 gene in TgME49 sequence). This indicates that both Ugandan strains are closer to each other than they are to the type II reference strain. As described in Chapter 2, Bontell et al. (2009) revealed that *in vitro* growth rate was higher in TgCkUg8 compared to the TgCkUg9 strain. Thus, the genetic variations between these two Ugandan strains within the virulence-associated gene ROP5 could be associated with variations in phenotypes of these strains.

The highly polymorphic ROP5 is a multicopy gene consisting of 4 to 10 copies which varies within different *Toxoplasma gondii* strains: ~6 copies in type I, ~10 in type II and ~4 in type III strains (Reese et al., 2011; Reese et al., 2014). In the current study, the variations have been identified through mapping the sequences of TgCkUg8 and 9 strains against the annotated sequences of the type II reference strain genome (TgME49) (retrieved from TOXODB) which has one sequence of ROP5 gene. The presence of a single sequence forROP5 in TgME49 genome can be explained by existence of multiple, closely identical genes within the ROP5 locus which had been viewed as a single sequence by the algorithm that used for genome assembly (Reese et al., 2011).

Variations between different ROP5 alleles that code for domains on n the surface of this protein were found to have an important role ine virulence through binding and interaction

with host's IRGs (Reese et al., 2014). Behnke et al. (2015) identified 23 codon sites under positive selection and majority of these polymorphic sites located in the surface of ROP5. By looking to the local variations in TgCkUg8 and 9 strains within ROP5 against type II and III reference strains (Figure 4.4), none of these amino acids are similar to the active 23 polymorphic surface sites mentioned earlier.

In this study, it was shown that detection of SNPs in the ROP5 gene in both Ugandan strains with 3 SNPs in TgCkUg8 and 8 polymorphisms in the TgCkUg9 strain evident that this virulence associated gene is a highly polymorphic candidate. This finding is in agreement with Bontell et al. (2009) study which showed that ROP5 is highly variable gene.

5. CHAPTER 5: General Discussion

In *Toxoplasma gondii*, several markers have been identified for the typing of isolates. Firstly, isoenzyme studies and RFLP methods were used before the development of sequencing of the genome of Toxoplasma. The availability of sequence information and the advances made by the Toxoplasma genome project have significantly improved identification of specific polymorphic markers. Application of multiple genetic markers in multi-locus analysis is necessary to estimate the parasite population genetics. Single nucleotide polymorphisms (SNPs) are considered to be the most common group of genetic markers, which were identified in single copy genes. They have been detected in genes coding for major antigens in rhoptries, dense granules and on the surface of the parasite. In addition, the SNPs have been found in genes coding structural proteins, for example in beta tubulin and enzymes, such as PK1, and in genes with unknown function, such as L358 and C29-2.

Analysis of single polymorphic genetic markers is insufficient, however, to detect recombination and to understand the population structure of the parasite. In this case, multi-locus genotyping is required to improve the sensitivity of the analysis through the number of polymorphic markers that are applied. For example, genotyping of recombinant I/III strains through application of six isoenzymes and five microsatellites revealed that these strains have 10 type I alleles and only 1 type III allele (Ajzenberg et al., 2002a), while these strains had been classified as zymodeme 1 (type I) in a previous study via analysis of only isoenzymes (Darde, 1996).

In SNP-RFLP markers, PCR-RFLP methods have been largely developed due to the detection of polymorphic restriction sites by using endonucleases. Although PCR-RFLP methods are cheaper and less time-consuming compared to DNA sequencing, they are less useful than sequencing methods. It is assumed that all Toxoplasma isolates follow a certain pattern of restriction sites which was defined mainly for strains collected from North America and Europe. Therefore, detecting SNPs in these markers by DNA sequencing can be directly applied as a typing method. In addition, multi-locus sequence typing (MLST) that utilize the number of loci is considered to be one of the best methods of identifying polymorphisms among isolates and analysis of the population structure of the parasite. For

instance, in a study conducted by (Fazaeli et al., 2000) it is shown that GRA6 sequence analysis of 30 stocks identified 9 allelic sequences - a high level of polymorphism compared to GRA6 PCR-RFLP technique which has the ability to differentiate only three groups of Toxoplasma strains.

In our study, it is shown that, when using the PCR RFLP method via 5 markers, it was very difficult to separate seven Ugandan isolates (Bontell et al., 2009), compared to our multi-locus PCR sequencing method using different nine markers to analyse the same set of isolates. In Chapter 2, it was reported that, by generating sequence data for these 9 loci in seven Ugandan isolates, the number of polymorphisms increased from 5, which represented the enzyme restriction sites, to 34 SNPs, which improved the discrimination power of these markers to detect the variation among the local Ugandan strains. This was sufficient to separate the additional 5 type II strains. Although most type II strains are differentiated by using the nine genetic markers, 4 of these strains are not distinguished from each other and are identical to the reference type II strain ME49 (TgCkUg1, 3, 8 and PRU). Thus, 9 out of 10 genetic markers (SAG1, SAG2, GRA6, BTUB, L358, C22-8, C29-2, PK1, and Apico except SAG3 which was not used) may be sufficient to separate global variation between *T.gondii* strains (Su et al., 2010; Su et al., 2012), but not to resolve within haplogroups (type II strains).

There is a step change between the MLST sequencing approach and the next available technology (genome sequencing) where thousands of loci may be examined. It is not appropriate to compare total genome data but deeper sequencing can be used to search for and identify polymorphic loci and generate a wider panel of markers. In Chapter 4, it was reported that new loci could be identified to probe phylogenetic relationships of closely related isolates. It was clear that six polymorphic markers (ROP7, ROP19A, ROP19B, ROP8, ROP5 and ROP on chr. XII), which were identified via analysis of three biologically relevant genes families (SRS, ROPs and GRA), enhanced the resolution power to identify variations among local type II strains of *T. gondii*. It is recommended that further study of these polymorphic markers is carried out and that they are added to the MLST analysis of *T. gondii*, especially between closely related local isolates.

The genetic recombination is influenced by two mechanisms that occur during meiosis, crossing-over and independent assortment. The markers that have the ability to detect the independent assortment, which is the separation of both parental homologous chromosomes,

should be found in different chromosomes, whilst the markers that are located on the same chromosome are required to identify the crossing-over, which represents genetic exchange of parental homologous chromosomes. There has been a long held view about the clonal propagation of *T. gondii*, suggesting that recombination rarely occurs, yet paradoxically experimental recombination has always suggested normal patterns of genetic exchange (Grigg et al., 2001a). The problem lies in the fact that MLST markers are very sparse across the genome, leaving very little opportunity to detect exchange. Markers are able to show global variation and resolve number of haplogroups, but not to look clearly and evaluate exchange within these haplogroups. Through the MLST sequencing approach, reported in Chapter 2, I found evidence of recombination type III regions within type II strains. The pattern of SNPs distribution in different loci across different chromosomes within three Ugandan strains (TgCkUg5, 6 and 7) was evidence of recombination between type II and III within these strains, to be added to a previously detected recombinant strain in Uganda, TgCkUg2, (Bontell et al., 2009). It was also evidenced by deeper sequencing of the TgCkUg9 strain (Chapter 3) that nine of fourteen chromosomes of this strain originated from a type II parent strain, suggesting recombination between type II and III. With the advent of a major new project sequencing 60 genomes of *T. gondii* strains representing all 15 major lineages of this parasite (ToxoDB) (Tables 3.5 and 3.6), this will allow a chromosome painting approach in which recombination events can be more readily identified.

Unbiased sampling of different isolates is required to understand the population structure for a pathogen (Darde, 2004). In *T. gondii*, it is difficult to meet this criterion as this parasite is widely distributed throughout a wide range of habitats and hosts. From the literature, it is shown that the majority of investigations and studies were performed on isolates from a relatively limited geographical range (North America and Europe). In addition, the majority of these isolates were collected from domestic animals (mainly chickens, sheep and pigs) and from humans with symptomatic toxoplasmosis (reactivation of infection in immunocompromised or congenital toxoplasmosis).

This bias is decreasing by extension of studies over a broader host and geographical range. Various studies have established typing analysis through the application of a monolocus typing (SAG2) on isolates collected from chickens in Brazil, other South American countries, Africa and Asia (Dubey et al., 2002; Dubey et al., 2003a; Dubey et al., 2003c; Dubey et al., 2004b; Dubey et al., 2005a; Dubey et al., 2005b), from Brazilian pigs, dogs,

cats and humans (da Silva et al., 2005; dos Santos et al., 2005; Dubey et al., 2004c; Vallochi et al., 2005) and from North American wildlife species (Dubey et al., 2004a). This geographical expansion in the collection of *T. gondii* samples will develop greater understanding of the genetic population structure of this parasite. However, these studies are still inadequate to understand the population structure as they were based only on one marker (monolocus typing – SAG2) which lacks the ability to identify genetic diversity between the isolates and recombination events. This issue can be improved through use of multi-locus typing analysis.

As sampling has become richer and more loci have been sequenced for a wider selection of strains, the pattern of clonal inheritance has been slowly breaking down. To really understand relationships and understand whether there is a panmictic structure to *T. gondii* populations, we need to have sufficient numbers of local isolates to be able to calculate allele frequencies. Only then will it be possible to work out what the relationships are. In this work we have perhaps the first example of deeper sequencing of isolates from the same geographical region at the same time point. It shows that they are non-identical and now that we have more potential markers we can use these to move forward. It is unrealistic to think that we will be able to investigate population structure by working on isolates. Lindstrom et al., (2008) showed that there is bias in the selection of isolates *in vivo* and *in vitro*; this means that we can never have 100% of infected tissue samples captured as isolates. In the future, the strategy moving forward will be to identify a richer panel of SNP's which can be used to expand MLST analysis. There is strong evidence that this approach will be productive. For example, in the 10 loci, 9 markers were used and they consistently worked. The next priority will be to work up additional markers on the basis of the sequencing strategy described above..

Intensive investigations that use polymorphic genetic markers are required to track different strains of *T. gondii* in order to identify and understand the different patterns of transmission of this parasite. It is shown that the genotype of a certain *T. gondii* isolate might suggest its geographical region more than reflecting its host species. For instance, infection of chickens in USA were found to have type I and III SAG2 strains of *T. gondii* (Dubey et al., 2003b), while the same species in Brazil carried mainly type II SAG2 strains (Dubey et al., 2002). Thus, biogeographical differences should be taken into consideration to understand the variations of transmission of *T. gondii* among hosts. There is not a single case of a

sympatric study of strains from different species. If, however, we have sufficient markers we can look at the profile of isolates and work out which species share the same parasites. This study shows for the first time that local strains are non-identical, lending credibility to this approach.

In the three main clonal lineages of *T. gondii*, virulence in mice is considered to be the most identified phenotypic marker. Virulence in mice is associated with the type I strain of the parasite, while the other two types, II and III, were suggested as avirulent strains in mice. Genetic diversity between these strains is associated with a variation in immune response of the host, which might help in understanding the variation in virulence patterns (Saeij et al., 2005). However, the host responses among different host species are not in the same pattern against virulent strains. For instance, the virulent type I strains in mice showed as a non-pathogenic feature in rats (Zenner et al., 1999). Virulence in mice of recombinant and atypical strains are generally greater compared with types II and III, and it is difficult to directly compare them while their genetic diversity is high. The variations in virulence in mice and other biological characteristics may be associated with variations in inherited genes combination in these strains (Grigg and Suzuki, 2003). Some members of these three families were seen to be associated with virulence in mice, including SAG1, SAG3, ROP16, ROP18 and ROP5 (Howe et al., 1996; Saeij et al., 2006; Taylor et al., 2006). *In vitro* studies of *T. gondii* biological characteristics have correlated with the studies in mice, suggesting greater virulence associated with type I strain compared with types II and III. These biological properties include: growth rate in cell culture, reinvasion rate, migration rate, penetration of submucosa and interconversion rate from tachyzoites to bradyzoites (Barragan and Sibley, 2002; Saeij et al., 2005).

In our study, (Chapter 4) novel polymorphisms were identified in a virulence-associated gene (ROP5) located on chromosome XII in both TgCkUg8 and 9 strains. These mutations, within the ROP5 gene resulted in modification of the protein structure by changing the composition of two amino acids in TgCkUg8, and four amino acids in the TgCkUg9 strain, compared with type II reference strain (TgME49). This genetic variation between TgCkUg8 and 9 strains within this virulence-associated gene (ROP5) could be associated with *in vitro* growth rate phenotype variation, which shown by Bontell et al. (2009) in these strains. In Chapter 2, it was also reported that the Bontell et al. (2009) study had found that three Ugandan strains (TgCkUg1, 3 and 8), which were genetically identical, had a higher growth

in vitro among eight Ugandan strains, providing evidence that genotype might influence phenotype within Ugandan strains.

Collection of isolates from human toxoplasmosis cases is rare. Samples collected for diagnosis are tachyzoites that have been isolated from pathological samples, such as amniotic fluid, blood, broncho-alveolar lavage, or from tissue cysts which are sampled by biopsy. Isolates of *T. gondii* from human clinical cases are mainly collected from immunocompromised patients or from congenital cases, while samples from immunocompetent cases with symptomatic toxoplasmosis are less common. It was suggested that strain isolation through mouse inoculation, even in clinical situations, may retrieve only one strain from a mixed infection via the effect of selective pressure (Villena et al., 2004).

Toxoplasma genotyping of primary clinical samples via direct analysis is theoretically possible due to the high sensitivity of PCR methods. In addition, the risk of bias in identification of strains that may be affected by isolation approaches will decrease, and genotyping could be applied on small amounts of clinical pathological samples such as ocular fluid (Grigg et al., 2001b). Even with the application of nested PCR, the very low number of *T. gondii* that exist in some clinical samples and the presence of PCR inhibitors can lead to negative results (Howe et al., 1997), in addition to the risk of DNA contamination by the use of nested PCR. It is clear that direct typing or strain isolation of pathological clinical samples were achieved only with symptomatic infection. Therefore, to understand the effect of different genotypes on clinical features of human toxoplasmosis, it is necessary to acquire genotype samples collected from asymptomatic human toxoplasmosis cases. It was revealed from early studies that type II strains were predominant in symptomatic toxoplasmosis patients from USA and Europe. This scenario might be different in other parts of the world.

The relationship of *T. gondii* genotype and human toxoplasmosis is present, but its identification is still difficult due to a shortage in genetic background and human immune response status (Suzuki et al., 1996). It is shown that type II strains of *T. gondii*, revealed to be the most common strain in human toxoplasmosis (Table 1.1), are virulent in some immunocompromised cases and immature foetuses. However, this strain was detected in several asymptomatic or mild infections in mature foetuses and was possibly associated with the majority of asymptomatic toxoplasmosis in immunocompetent cases in Europe. This

may be the same in type III infections, although studies that indicate its association with human toxoplasmosis are few. Some toxoplasmosis cases have revealed the association of type I strains with a greater virulence level, such as acquired ocular infection and disseminated congenital cases (Fuentes et al., 2001; Vallochi et al., 2005). However, its association with asymptomatic toxoplasmosis in immunocompromised cases was suggested because of the detection of this strain in some placentas without congenital manifestations, and in some cases among immunocompromised placentas due to the reactivation of chronic toxoplasmosis. In recombinant and atypical strains, the condition is more complicated, depending on the combination of genes of different *T. gondii* types. It has been proposed that virulence in recombinant strains is associated with the existence of type I alleles (Carme et al., 2002; Grigg et al., 2001b). Thus, to understand the genetic basis that contributes virulence in human toxoplasmosis with recombinant strains, further studies of these strains in humans accompanied by experimental investigations in mice of experimental and natural recombinants are required in future.

Human associations are particularly hard to find, as phenotype is hard to capture. A different approach is needed to ensure that we can relate phenotype to genotype; this is an increased ability to type from tissue samples, plus new ways of capturing disease phenotype. As a stepping stone we could do more to set standards for the capture of virulence *in vivo*, through monitoring parasite density or immunological responses.

In conclusion, we have come a long way in understanding this ubiquitous parasite. Gradually, we are seeing that recombination may not be an unusual event and so the priority over the next few years must be to work much more closely on levels of local variation to genotype strains in sympatric studies of species in relation to phenotype. The great asset of this system is that levels of genetic variation are low so identification of virulence associated loci should be possible. It may be timely to return to Uganda with the aim of launching such a study.

REFERENCES

Ajzenberg, D., Banuls, A. L., Su, C., Dumetre, A., Demar, M., Carme, B. & Darde, M. L. (2004). Genetic diversity, clonality and sexuality in *Toxoplasma gondii*. *International Journal for Parasitology,* 34**,** 1185-1196.

Ajzenberg, D., Banuls, A. L., Tibayrenc, M. & Darde, M. L. (2002a). Microsatellite analysis of *Toxoplasma gondii* shows considerable polymorphism structured into two main clonal groups. *International Journal for Parasitology,* 32**,** 27-38.

Ajzenberg, D., Cogne, N., Paris, L., Bessieres, M. H., Thulliez, P., Filisetti, D., Pelloux, H., Marty, P. & Darde, M. L. (2002b). Genotype of 86 *Toxoplasma gondii* isolates associated with human congenital toxoplasmosis, and correlation with clinical findings. *Journal of Infectious Diseases,* 186**,** 684-689.

Asai, T., Miura, S., Sibley, L. D., Okabayashi, H. & Takeuchi, T. (1995). Biochemical and molecular characterization of nucleoside triphosphate hydrolase isozymes from the parasitic protozoan *Toxoplasma gondii*. *Journal of Biological Chemistry,* 270**,** 11391-11397.

Aspinall, T. V., Joynson, D. H. M., Guy, E., Hyde, J. E. & Sims, P. F. G. (2002). The molecular basis of sulfonamide resistance in *Toxoplasma gondii* and implications for the clinical management of toxoplasmosis. *Journal of Infectious Diseases,* 185**,** 1637-1643.

Banuls, A. L., Hide, M. & Tibayrenc, M. (1999). Molecular epidemiology and evolutionary genetics of Leishmania parasites. *International Journal for Parasitology,* 29**,** 1137-1147.

Barragan, A. & Sibley, L. D. (2002). Transepithelial migration of *Toxoplasma gondii* is linked to parasite motility and virulence. *Journal of Experimental Medicine,* 195**,** 1625-1633.

Barragan, A. & Sibley, L. D. (2003). Migration of *Toxoplasma gondii* across biological barriers. *Trends in Microbiology,* 11**,** 426-430.

Barry, J., Marcello, L., Morrison, L., Read, A., Lythgoe, K., Jones, N., Carrington, M., Blandin, G., Bohme, U. & Caler, E. (2005). What the genome sequence is revealing about trypanosome antigenic variation. *Biochemical Society Transactions,* 33**,** 986-989.

Behnke, M. S., Khan, A., Lauron, E. J., Jimah, J. R., Wang, Q., Tolia, N. H. & Sibley, L. D. (2015). Rhoptry proteins ROP5 and ROP18 are major murine virulence factors in genetically divergent South American strains of *Toxoplasma gondii*. *PLoS Genetics,* 11, e1005434.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, W., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L. H., Wickstead, B., Alsmark, U. C. M., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetter, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M. A., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Talton, L., Turner, C. M. R., Tait, A., Tivey, A. R.,

Van Aken, S., Walker, D., Wanless, D., Wang, S. L., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barret, B. G., Donelson, J. E., Hall, N., Fraser, C. M., et al. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science,* 309**,** 416-422.

Besteiro, S., Dubremetz, J. F. & Lebrun, M. (2011). The moving junction of apicomplexan parasites: a key structure for invasion. *Cellular Microbiology,* 13**,** 797-805.

Beverley, J. K. A. (1959). Congenital transmission of toxoplasmosis through successive generations of mice. *Nature,* 183**,** 1348-1349.

Bezerra, R. A., Carvalho, F. S., Guimaraes, L. A., Rocha, D. S., Maciel, B. M., Wenceslau, A. A., Lopes, C. W. G. & Albuquerque, G. R. (2012). Genetic characterization of *Toxoplasma gondii* isolates from pigs intended for human consumption in Brazil. *Veterinary Parasitology,* 189**,** 153-161.

Binas, M. & Johnson, A. M. (1998). A polymorphism in a DNA polymerase alpha gene intron differentiates between murine virulent and avirulent strains of *Toxoplasma gondii*. *International Journal for Parasitology,* 28**,** 1033-1040.

Bontell, I. L., Hall, N., Ashelford, K. E., Dubey, J. P., Boyle, J. P., Lindh, J. & Smith, J. E. (2009). Whole genome sequencing of a natural recombinant *Toxoplasma gondii* strain reveals chromosome sorting and local allelic variants. *Genome Biology,* 10, R53

Boothroyd, J. C. & Dubremetz, J. F. (2008). Kiss and spit: the dual roles of Toxoplasma rhoptries. *Nature Reviews Microbiology,* 6**,** 79-88.

Boothroyd, J. C. & Grigg, M. E. (2002). Population biology of *Toxoplasma gondii* and its relevance to human infection: do different strains cause different disease? *Current Opinion in Microbiology,* 5**,** 438-442.

Boothroyd, J. C., Hehl, A., Knoll, L. J. & Manger, I. D. (1998). The surface of Toxoplasma: more and less. *International Journal for Parasitology,* 28**,** 3-9.

Boudeau, J., Miranda-Saavedra, D., Barton, G. J. & Alessi, D. R. (2006). Emerging roles of pseudokinases. *Trends in Cell Biology,* 16**,** 443-452.

Boyle, J. P., Rajasekar, B., Saeij, J. P. J., Ajiokat, J. W., Berriman, M., Paulsen, I., Roos, D. S., Sibley, L. D., White, M. W. & Boothroyd, J. C. (2006). Just one cross appears capable of dramatically altering the population biology of a eukaryotic pathogen like *Toxoplasma gondii*. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 10514-10519.

Bradley, P. J., Ward, C., Cheng, S. J., Alexander, D. L., Coller, S., Coombs, G. H., Dunn, J. D., Ferguson, D. J., Sanderson, S. J., Wastling, J. M. & Boothroyd, J. C. (2005). Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in *Toxoplasma gondii*. *Journal of Biological Chemistry,* 280**,** 34245-34258.

Brayton, K. A., Lau, A. O. T., Herndon, D. R., Hannick, L., Kappmeyer, L. S., Berens, S. J., Bidwell, S. L., Brown, W. C., Crabtree, J., Fadrosh, D., Feldblum, T., Forberger, H. A., Haas, B. J., Howell, J. M., Khouri, H., Koo, H., Mann, D. J., Norimine, J., Paulsen, I. T., Radune, D., Ren, Q. H., Smith, R. K., Suarez, C. E., White, O., Wortman, J. R., Knowles, D. P., Mcelwain, T. F. & Nene, V. M. (2007). Genome sequence of babesia bovis and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathogens,* 3**,** 1401-1413.

Brezin, A. P., Kasner, L., Thulliez, P., Li, Q., Daffos, F., Nussenblatt, R. B. & Chan, C. C. (1994). Ocular toxoplasmosis in the fetus immunohistochemistry analysis and DNA amplification. *Retina-the Journal of Retinal and Vitreous Diseases,* 14**,** 19-26.

Burg, J. L., Perelman, D., Kasper, L. H., Ware, P. L. & Boothroyd, J. C. (1988). Molecular analysis of the gene encoding the major surface antigen of *Toxoplasma gondii*. *Journal of Immunology,* 141**,** 3584-3591.

Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C. M., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van De Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., Muller, S., Dessi, D., Fiori, P. L., Ren, Q. H., Paulsen, I., Zhang, H. B., Bastida-Corcuera, F. D., Simoes-Barbosa, A., Brown, M. T., Hayes, R. D., Mukherjee, M., Okumura, C. Y., Schneider, R., Smith, A. J., Vanacova, S., Villalvazo, M., Haas, B. J., Pertea, M., Feldblyum, T. V., Utterback, T. R., Shu, C. L., Osoegawa, K., De Jong, P. J., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S. B., Logsdon, J. M., Henze, K., Gupta, A., Wang, C. C., Dunne, R. L., Upcroft, J. A., Upcroft, P., White, O., Salzberg, S. L., Tang, P., Chiu, C. H., Lee, Y. S., Embley, T. M., Coombs, G. H., Mottram, J. C., Tachezy, J., Fraser-Liggett, C. M. & Johnson, P. J. (2007). Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science,* 315**,** 207-212.

Carme, B., Bissuel, F., Ajzenberg, D., Bouyne, R., Aznar, C., Demar, M., Bichat, S., Louvel, D., Bourbigot, A. M., Peneau, C., Neron, P. & Darde, M. L. (2002). Severe acquired toxoplasmosis in immunocompetent adult patients in French Guiana. *Journal of Clinical Microbiology,* 40**,** 4037-4044.

Chen, Z. W., Gao, J. M., Huo, X. X., Wang, L., Yu, L., Halm-Lai, F., Xu, Y. H., Song, W. J., Hide, G., Shen, J. L. & Lun, Z. R. (2011). Genotyping of *Toxoplasma gondii* isolates from cats in different geographic regions of China. *Veterinary Parasitology,* 183**,** 166-170.

Conrad, P. A., Miller, M. A., Kreuder, C., James, E. R., Mazet, J., Dabritz, H., Jessup, D. A., Gulland, F. & Grigg, M. E. (2005). Transmission of Toxoplasma: Clues from the study of sea otters as sentinels of *Toxoplasma gondii* flow into the marine environment. *International Journal for Parasitology,* 35**,** 1155-1168.

Cooper, J. E. & Feil, E. J. (2006). The phylogeny of *Staphylococcus aureus* - which genes make the best intra-species markers? *Microbiology,* 152**,** 1297-1305.

Coppens, I., Dunn, J. D., Romano, J. D., Pypaert, M., Zhang, H., Boothroyd, J. C. & Joiner, K. A. (2006). *Toxoplasma gondii* sequesters lysosomes from mammalian hosts in the vacuolar space. *Cell,* 125**,** 261-274.

Costa, J. M., Darde, M. L., Assouline, B., Vidaud, M. & Bretagne, S. (1997). Microsatellite in the beta-tubulin gene of *Toxoplasma gondii* as a new genetic marker for use in direct screening of amniotic fluids. *Journal of Clinical Microbiology,* 35**,** 2542-2545.

Craver, M. P. J. & Knoll, L. J. (2007). Increased efficiency of homologous recombination in *Toxoplasma gondii* dense granule protein 3 demonstrates that GRA3 is not necessary in cell culture but does contribute to virulence. *Molecular and Biochemical Parasitology,* 153**,** 149-157.

Da Silva, A. V., Pezerico, S. B., De Lim, V. Y., Moretti, L. D., Pinheiro, J. P., Tanaka, E. M., Ribeiro, M. G. & Langoni, H. (2005). Genotyping of *Toxoplasma gondii* strains isolated from dogs with neurological signs. *Veterinary Parasitology,* 127**,** 23-27.

Dalmasso, M. C., Onyango, D. O., Naguleswaran, A., Sullivan, W. J. & Angel, S. O. (2009). Toxoplasma H2A Variants Reveal Novel Insights into Nucleosome Composition and Functions for this Histone Family. *Journal of Molecular Biology,* 392**,** 33-47.

Dao, A., Fortier, B., Soete, M., Plenat, F. & Dubremetz, J. F. (2001). Successful reinfection of chronically infected mice by a different *Toxoplasma gondii* genotype. *International Journal for Parasitology,* 31**,** 63-65.

Darde, M. L. (2004). Genetic analysis of the diversity in *Toxoplasma gondii*. *Annali dell'Istituto superiore di sanita,* 40**,** 57-63.

Darde, M. L. (1996). Biodiversity in *Toxoplasma gondii*. *Current Topics in Microbiology and Immunology,* 219**,** 27-41.

Darde, M. L. (2008). *Toxoplasma gondii*, "new" genotypes and virulence. *Parasite-Journal De La Societe Francaise De Parasitologie,* 15**,** 366-371.

Darde, M. L., Bouteille, B. & Pestrealexandre, M. (1988). Isoenzymic characterization of 7 strains of *Toxoplasma gondii* by isoelectrofocusing in polyacrylamide gels. *American Journal of Tropical Medicine and Hygiene,* 39**,** 551-558.

Darde, M. L., Bouteille, B. & Pestrealexandre, M. (1992). Isoenzyme analysis of 35 *Toxoplasma gondii* isolates and the biological and epidemiologic implications. *Journal of Parasitology,* 78**,** 786-794.

De Sousa, S., Ajzenberg, D., Canada, N., Freire, L., Da Costa, J. M. C., Darde, M. L., Thulliez, P. & Dubey, J. P. (2006). Biologic and molecular characterization of *Toxoplasma gondii* isolates from pigs from Portugal. *Veterinary Parasitology,* 135**,** 133-136.

Delorme-Walker, V., Abrivard, M., Lagal, V., Anderson, K., Perazzi, A., Gonzalez, V., Page, C., Chauvet, J., Ochoa, W., Volkmann, N., Hanein, D. & Tardieux, I. (2012). Toxofilin upregulates the host cortical actin cytoskeleton dynamics, facilitating Toxoplasma invasion. *Journal of Cell Science,* 125**,** 4333-4342.

Denkers, E. Y. & Gazzinelli, R. T. (1998). Regulation and function of T-cell-mediated immunity during *Toxoplasma gondii* infection. *Clinical Microbiology Reviews,* 11**,** 569-588.

Didelot, X. & Maiden, M. C. J. (2010). Impact of recombination on bacterial evolution. *Trends in Microbiology,* 18**,** 315-322.

Dos Santos, C. B. D., De Carvalho, A., Ragozo, A. M. A., Soares, R. M., Amaku, M., Yai, L. E. O., Dubey, J. P. & Gennari, S. M. (2005). First isolation and molecular characterization of *Toxoplasma gondii* from finishing pigs from Sao Paulo State, Brazil. *Veterinary Parasitology,* 131**,** 207-211.

Dubey, J., Graham, D., Blackston, C., Lehmann, T., Gennari, S., Ragozo, A., Nishi, S., Shen, S., Kwok, O. & Hill, D. (2002). Biological and genetic characterisation of *Toxoplasma gondii* isolates from chickens (Gallus domesticus) from Sao Paulo, Brazil: unexpected findings. *International Journal for Parasitology,* 32**,** 99-105.

Dubey, J. P. (1980). Mouse pathogenicity of *Toxoplasma gondii* isolated from a goat. *American Journal of Veterinary Research,* 41**,** 427-429.

Dubey, J. P. (1997). Tissue cyst tropism in *Toxoplasma gondii*: a comparison of tissue cyst formation in organs of cats, and rodents fed oocysts. *Parasitology,* 115**,** 15-20.

Dubey, J. P. (1998a). Advances in the life cycle of *Toxoplasma gondii*. *International Journal for Parasitology,* 28**,** 1019-1024.

Dubey, J. P. (1998b). Bradyzoite-induced murine toxoplasmosis: Stage conversion, pathogenesis, and tissue cyst formation in mice fed bradyzoites of different strains of *Toxoplasma gondii*. *Journal of Eukaryotic Microbiology,* 44**,** 592-602.

Dubey, J. P. (2004). Toxoplasmosis - a waterborne zoonosis. *Veterinary Parasitology,* 126**,** 57-72.

Dubey, J. P. (2006). Comparative infectivity of oocysts and bradyzoites of *Toxoplasma gondii* for intermediate (mice) and definitive (cats) hosts. *Veterinary Parasitology,* 140**,** 69-75.

Dubey, J. P. (2007). The history and life cycle of *Toxoplasma gondii*. *In:* Louis, M. W. & Kami, K. (Eds.) *Toxoplasma gondii.* London: Academic Press.

Dubey, J. P. (2009a). History of the discovery of the life cycle of *Toxoplasma gondii*. *International Journal for Parasitology,* 39**,** 877-882.

Dubey, J. P. (2009b). *Toxoplasmosis of Animals and Humans, Second Edition*, CRC Press.

Dubey, J. P. & Frenkel, J. K. (1972). Cyst-induced toxoplasmosis in cats. *Journal of Protozoology,* 19**,** 155-&.

Dubey, J. P., Graham, D. H., Dahl, E., Hilali, M., El-Ghaysh, A., Sreekumar, C., Kwok, O. C. H., Shen, S. K. & Lehmann, T. (2003a). Isolation and molecular characterization of *Toxoplasma gondii* from chickens and ducks from Egypt. *Veterinary Parasitology,* 114**,** 89-95.

Dubey, J. P., Graham, D. H., Dahl, E., Sreekumar, C., Lehmann, T., Davis, M. F. & Morishita, T. Y. (2003b). *Toxoplasma gondii* isolates from free-ranging chickens from the United States. *Journal of Parasitology,* 89**,** 1060-1062.

Dubey, J. P., Graham, D. H., De Young, R. W., Dahl, E., Eberhard, M. L., Nace, E. K., Won, K., Bishop, H., Punkosdy, G., Sreekumar, C., Vianna, M. C. B., Shen, S. K., Kwok, O. C. H., Sumners, J. A., Demarais, S., Humphreys, J. G. & Lehmann, T. (2004a). Molecular and biologic characteristics of *Toxoplasma gondii* isolates from wildlife in the United States. *Journal of Parasitology,* 90**,** 67-71.

Dubey, J. P. & Jones, J. L. (2008). *Toxoplasma gondii* infection in humans and animals in the United States. *International Journal for Parasitology,* 38**,** 1257-1278.

Dubey, J. P., Karhemere, S., Dahl, E., Sreekumar, C., Diabate, A., Dabire, K. R., Vianna, M. C. B., Kwok, O. C. H. & Lehmann, T. (2005a). First biologic and genetic characterization of *Toxoplasma gondii* isolates from chickens from Africa (Democratic Republic of Congo, Mali, Burkina Faso, and Kenya). *Journal of Parasitology,* 91**,** 69-72.

Dubey, J. P., Levy, M. Z., Sreekumar, C., Kwok, O. C. H., Shen, S. K., Dahl, E., Thulliez, P. & Lehmann, T. (2004b). Tissue distribution and molecular characterization of chicken isolates of *Toxoplasma gondii* from Peru. *Journal of Parasitology,* 90**,** 1015-1018.

Dubey, J. P., Lindsay, D. S. & Speer, C. A. (1998). Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. *Clinical Microbiology Reviews,* 11**,** 267-299.

Dubey, J. P., Navarro, I. T., Sreekumar, C., Dahl, E., Freire, R. L., Kawabata, H. H., Vianna, M. C. B., Kwok, O. C. H., Shen, S. K., Thulliez, P. & Lehmann, T. (2004c). *Toxoplasma gondii* infections in cats from Parana, Brazil: Seroprevalence, tissue distribution, and biologic and genetic characterization of isolates. *Journal of Parasitology,* 90**,** 721-726.

Dubey, J. P., Rajapakse, R., Ekanayake, D. K., Sreekumar, C. & Lehmann, T. (2005b). Isolation and molecular characterization of *Toxoplasma gondii* from chickens from Sri Lanka. *Journal of Parasitology,* 91**,** 1480-1482.

Dubey, J. P., Velmurugan, G. V., Chockalingam, A., Pena, H. F. J., De Oliveira, L. N., Leifer, C. A., Gennari, S. M., Oliveira, L. & Su, C. (2008). Genetic diversity of *Toxoplasma gondii* isolates from chickens from Brazil. *Veterinary Parasitology,* 157**,** 299-305.

Dubey, J. P., Venturini, M. C., Venturini, L., Piscopo, M., Graham, D. H., Dahl, E., Sreekumar, C., Vianna, M. C. & Lehmann, T. (2003c). Isolation and genotyping of *Toxoplasma gondii* from free-ranging chickens from Argentina. *Journal of Parasitology,* 89**,** 1063-1064.

Dubremetz, J. F., Garcia-Reguet, N., Conseil, V. & Fourmaux, M. N. (1998). Apical organelles and host-cell invasion by Apicomplexa. *International Journal for Parasitology,* 28**,** 1007-1013.

Dubremetz, J. F. & Lebrun, M. (2012). Virulence factors of *Toxoplasma gondii*. *Microbes and Infection,* 14**,** 1403-1410.

Dunn, D., Wallon, M., Peyron, F., Petersen, E., Peckham, C. & Gilbert, R. (1999). Mother-to-child transmission of toxoplasmosis: risk estimates for clinical counselling. *Lancet,* 353**,** 1829-1833.

Dzierszinski, F., Mortuaire, M., Cesbron-Delauw, M. F. & Tomavo, S. (2000). Targeted disruption of the glycosylphosphatidylinositol-anchored surface antigen SAG3 gene in *Toxoplasma gondii* decreases host cell adhesion and drastically reduces virulence in mice. *Molecular Microbiology,* 37**,** 574-582.

El Hajj, H., Demey, E., Poncet, J., Lebrun, M., Wu, B., Galéotti, N., Fourmaux, M. N., Mercereau-Puijalon, O., Vial, H. & Labesse, G. (2006). The ROP2 family of *Toxoplasma gondii* rhoptry proteins: proteomic and genomic characterization and molecular modeling. *Proteomics,* 6**,** 5773-5784.

El Hajj, H., Lebrun, M., Arold, S. T., Vial, H., Labesse, G. & Dubremetz, J. F. (2007). ROP18 is a rhoptry kinase controlling the intracellular proliferation of *Toxoplasma gondii*. *PLoS Pathogens,* 3**,** 200-211.

El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., Da Silveira, J. F., De Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L. H., Huang, Y. T., Kindlund, E., Ktingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., Mcculloch, R., Mckenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarteton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D. & Andersson, B. (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science,* 309**,** 409-415.

Falush, D., Stephens, M. & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics,* 164**,** 1567-1587.

Fazaeli, A., Carter, P. E., Darde, M. L. & Pennington, T. H. (2000). Molecular typing of *Toxoplasma gondii* strains by GRA6 gene sequence analysis. *International Journal for Parasitology,* 30**,** 637-642.

Feil, E. J., Maiden, M. C. J., Achtman, M. & Spratt, B. G. (1999). The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Molecular Biology and Evolution,* 16**,** 1496-1502.

Fentress, S. J., Behnke, M. S., Dunay, I. R., Mashayekhi, M., Rommereim, L. M., Fox, B. A., Bzik, D. J., Taylor, G. A., Turk, B. E., Lichti, C. F., Townsend, R. R., Qiu, W., Hui, R., Beatty, W. L. & Sibley, L. D. (2010). Phosphorylation of immunity-related GTPases by a *Toxoplasma gondii*-secreted sinase promotes macrophage survival and virulence. *Cell Host & Microbe,* 8**,** 484-495.

Ferreira, A. D., Vitor, R. W. A., Gazzinelli, R. T. & Melo, M. N. (2006). Genetic analysis of natural recombinant Brazilian *Toxoplasma gondii* strains by multilocus PCR-RFLP. *Infection Genetics and Evolution,* 6**,** 22-31.

Field, D. & Wills, C. (1996). Long, polymorphic microsatellites in simple organisms. *Proceedings of the Royal Society B-Biological Sciences,* 263**,** 209-215.

Forrester, S. J. & Hall, N. (2014). The revolution of whole genome sequencing to study parasites. *Molecular and Biochemical Parasitology,* 195**,** 77-81.

Frazao-Teixeira, E., Sundar, N., Dubey, J. P., Grigg, M. E. & De Oliveira, F. C. R. (2011). Multi-locus DNA sequencing of *Toxoplasma gondii* isolated from Brazilian pigs identifies genetically divergent strains. *Veterinary Parasitology,* 175**,** 33-39.

Frenkel, J. K. (1973). Toxoplasma in and around United States. *Bioscience,* 23**,** 343-353.

Fuentes, I., Rubio, J. M., Ramirez, C. & Alvar, J. (2001). Genotypic characterization of *Toxoplasma gondii* strains associated with human toxoplasmosis in Spain: Direct analysis from clinical samples. *Journal of Clinical Microbiology,* 39**,** 1566-1570.

Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J. C., Mackey, A. J., Pinney, D. F., Roos, D. S., Stoeckert, C. J., Wang, H. & Brunk, B. P. (2008). ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Research,* 36**,** D553-D556.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M. A., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., Mcfadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M. & Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature,* 419**,** 498-511.

Gazzinelli, R. T., Wysocka, M., Hayashi, S., Denkers, E. Y., Hieny, S., Caspar, P., Trinchieri, G. & Sher, A. (1994). Parasite-induced IL-12 stimulates early IFN-gamma synthesis and resistance during acute infection with *Toxoplasma gondii*. *Journal of Immunology,* 153**,** 2533-2543.

Gilbert, L. A., Ravindran, S., Turetzky, J. M., Boothroyd, J. C. & Bradley, P. J. (2007). *Toxoplasma gondii* targets a protein phosphatase 2C to the nuclei of infected host cells. *Eukaryotic cell,* 6**,** 73-83.

Gissot, M., Kelly, K. A., Ajioka, J. W., Greally, J. M. & Kim, K. (2007). Epigenomic modifications predict active promoters and gene structure in *Toxoplasma gondii*. *PLoS Pathogens,* 3**,** 709-719.

Gissot, M., Walker, R., Delhaye, S., Huot, L., Hot, D. & Tomavo, S. (2012). *Toxoplasma gondii* chromodomain protein 1 binds to heterochromatin and colocalises with centromeres and telomeres at the nuclear periphery. *PLoS One,* 7, e32671.

Goodhead, I., Capewell, P., Bailey, J. W., Beament, T., Chance, M., Kay, S., Forrester, S., Macleod, A., Taylor, M., Noyes, H. & Hall, N. (2013). Whole-genome sequencing of *Trypanosoma brucei* reveals introgression between subspecies that is associated with virulence. *Mbio,* 4, e00197-13.

Grigg, M. E., Bonnefoy, S., Hehl, A. B., Suzuki, Y. & Boothroyd, J. C. (2001a). Success and virulence in Toxoplasma as the result of sexual recombination between two distinct ancestries. *Science,* 294**,** 161-165.

Grigg, M. E., Ganatra, J., Boothroyd, J. C. & Margolis, T. P. (2001b). Unusual abundance of atypical strains associated with human ocular toxoplasmosis. *Journal of Infectious Diseases,* 184**,** 633-639.

Grigg, M. E. & Suzuki, Y. (2003). Sexual recombination and clonal evolution of virulence in Toxoplasma. *Microbes and Infection,* 5**,** 685-690.

Grimwood, J. & Smith, J. E. (1992). *Toxoplasma gondii* the role of a 30-kda surface protein in host cell invasion. *Experimental Parasitology,* 74**,** 106-111.

Grimwood, J. & Smith, J. E. (1996). *Toxoplasma gondii*: The role of parasite surface and secreted proteins in host cell invasion. *International Journal for Parasitology,* 26**,** 169-173.

Heinz, E., Williams, T. A., Nakjang, S., Noel, C. J., Swan, D. C., Goldberg, A. V., Harris, S. R., Weinmaier, T., Markert, S., Becher, D., Bernhardt, J., Dagan, T., Hacker, C., Lucocq, J. M., Schweder, T., Rattei, T., Hall, N., Hirt, R. P. & Embley, T. M. (2012). The genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into Microsporidian genome dynamics and reductive evolution. *PLoS Pathogens,* 8, e1002979.

Hide, G., Morley, E. K., Hughes, J. M., Gerwash, O., Elmahaishi, M. S., Elmahaishi, K. H., Thomasson, D., Wright, E. A., Williams, R. H., Murphy, R. G. & Smith, J. E. (2009). Evidence for high levels of vertical transmission in *Toxoplasma gondii*. *Parasitology,* 136**,** 1877-1885.

Hill, D. & Dubey, J. P. (2002). *Toxoplasma gondii*: transmission, diagnosis and prevention. *Clinical Microbiology and Infection,* 8**,** 634-640.

Ho-Yen, D. O. (2009). Toxoplasmosis. *Medicine,* 37**,** 665-667.

Holland, G. N. (2003). Ocular toxoplasmosis: a global reassessment - Part I: epidemiology and course of disease. *American Journal of Ophthalmology,* 136**,** 973-988.

Holland, G. N. (2004). Ocular toxoplasmosis: a global reassessment - Part II: disease manifestations and management. *American Journal of Ophthalmology,* 137**,** 1-17.

Holland, G. N., Engstrom, R. E., Glasgow, B. J., Berger, B. B., Daniels, S. A., Sidikaro, Y., Harmon, J. A., Fischer, D. H., Boyer, D. S. & Rao, N. A. (1988). Ocular toxoplasmosis in patients with the acquired immunodeficiency syndrome. *American Journal of Ophthalmology,* 106**,** 653-667.

Howe, D. K., Honore, S., Derouin, F. & Sibley, L. D. (1997). Determination of genotypes of *Toxoplasma gondii* strains isolated from patients with toxoplasmosis. *Journal of Clinical Microbiology,* 35**,** 1411-1414.

Howe, D. K. & Sibley, L. D. (1995). *Toxoplasma gondii* comprises 3 clonal lineages correlation of parasite genotype with human disease. *Journal of Infectious Diseases,* 172**,** 1561-1566.

Howe, D. K., Summers, B. C. & Sibley, L. D. (1996). Acute virulence in mice is associated with markers on chromosome VIII in *Toxoplasma gondii*. *Infection and Immunity,* 64**,** 5193-5198.

Hoyen, D. O. (1990). Toxoplasmosis in humans - discussion paper. *Journal of the Royal Society of Medicine,* 83**,** 571-572.

Hutchison,W. M. (1965). Experimental transmission of *Toxoplasma gondii*. *Nature,* 206**,** 961-962.

Hutchison, W. M., Dunachie, J. F., Work, K. & Chr. Siim, J. (1971). The life cycle of the coccidian parasite, *Toxoplasma gondii*, in the domestic cat. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 65**,** 380-398.

Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchettin, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M. R., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y. T., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony,

M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S. G., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B. & Myler, P. J. (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science,* 309**,** 436-442.

Jackson, A. P., Berry, A., Aslett, M., Allison, H. C., Burton, P., Vavrova-Anderson, J., Brown, R., Browne, H., Corton, N., Hauser, H., Gamble, J., Gilderthorp, R., Marcello, L., Mcquillan, J., Otto, T. D., Quail, M. A., Sanders, M. J., Van Tonder, A., Ginger, M. L., Field, M. C., Barry, J. D., Hertz-Fowler, C. & Berriman, M. (2012). Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 3416-3421.

Jerome, M. E., Radke, J. R., Bohne, W., Roos, D. S. & White, M. W. (1998). *Toxoplasma gondii* bradyzoites form spontaneously during sporozoite-initiated development? *Infection and Immunity,* 66**,** 4838-4844.

Jung, C., Lee, C. Y. F. & Grigg, M. E. (2004). The SRS superfamily of Toxoplasma surface proteins. *International Journal for Parasitology,* 34**,** 285-296.

Kashi, Y., King, D. & Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics,* 13**,** 74-78.

Kasper, L. H., Crabb, J. H. & Pfefferkorn, E. R. (1983). Purification of a major membrane protein of *Toxoplasma gondii* by immunoabsorption with a monoclonal antibody. *Journal of Immunology,* 130**,** 2407-2412.

Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J. & Vivares, C. P. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature,* 414**,** 450-453.

Kaufman, H. E., Melton, M. L., Remington, J. S. & Jacobs, L. (1959). Strain differences of *Toxoplasma gondii*. *Journal of Parasitology,* 45**,** 189-190.

Kaufman, H. E., Remington, J. S. & Jacobs, L. (1958). Toxoplasmosis - the nature of virulence. *American Journal of Ophthalmology,* 46**,** 255-261.

Khan, A., Bohme, U., Kelly, K. A., Adlem, E., Brooks, K., Simmonds, M., Mungall, K., Quail, M. A., Arrowsmith, C., Chillingworth, T., Churcher, C., Harris, D., Collins, M., Fosker, N., Fraser, A., Hance, Z., Jagels, K., Moule, S., Murphy, L., O'Neil, S., Rajandream, M. A., Saunders, D., Seeger, K., Whitehead, S., Mayr, T., Xuan, X. N., Watanabe, J., Suzuki, Y., Wakaguri, H., Sugano, S., Sugimoto, C., Paulsen, I., Mackey, A. J., Roos, D. S., Hall, N., Berriman, M., Barrell, B., Sibley, L. D. & Ajioka, J. W. (2006a). Common inheritance of chromosome Ia associated with clonal expansion of *Toxoplasma gondii*. *Genome Research,* 16**,** 1119-1125.

Khan, A., Dubey, J., Su, C., Ajioka, J. W., Rosenthal, B. M. & Sibley, L. D. (2011). Genetic analyses of atypical *Toxoplasma gondii* strains reveal a fourth clonal lineage in North America. *International journal for parasitology,* 41**,** 645-655.

Khan, A., Fux, B., Su, C., Dubey, J. P., Darde, M. L., Ajioka, J. W., Rosenthal, B. M. & Sibley, L. D. (2007). Recent transcontinental sweeps of *Toxoplasma gondii* driven by a single monomorphic chromosome. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 14872-14877.

Khan, A., Jordan, C., Muccioli, C., Vallochi, A. L., Rizzo, L. V., Belfort, R., Vitor, R. W. A., Silveira, C. & Sibley, L. D. (2006b). Genetic divergence of *Toxoplasma gondii*

strains associated with ocular toxoplasmosis, Brazil. *Emerging Infectious Diseases,* 12**,** 942-949.

Khan, A., Taylor, S., Su, C., Mackey, A. J., Boyle, J., Cole, R., Glover, D., Tang, K., Paulsen, I. T., Berriman, M., Boothroyd, J. C., Pfefferkorn, E. R., Dubey, J. P., Ajioka, J. W., Roos, D. S., Wootton, J. C. & Sibley, L. D. (2005). Composite genome map and recombination parameters derived from three archetypal lineages of *Toxoplasma gondii*. *Nucleic Acids Research,* 33**,** 2980-2992.

Kim, K. & Weiss, L. M. (2004). *Toxoplasma gondii*: the model apicomplexan. *International Journal for Parasitology,* 34**,** 423-432.

Kim, S. K. & Boothroyd, J. C. (2005). Stage specific expression of surface antigens by *Toxoplasma gondii* as a mechanism to facilitate parasite persistence. *Journal of Immunology,* 174**,** 8038-8048.

Koren, S., Harhay, G. P., Smith, T. P., Bono, J. L., Harhay, D. M., Mcvey, S. D., Radune, D., Bergman, N. H. & Phillippy, A. M. (2013). Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biology,* 14, R101.

Kravetz, J. D. & Federman, D. G. (2005). Toxoplasmosis in pregnancy. *American Journal of Medicine,* 118**,** 212-216.

Kuhn, G., Francioli, P. & Blanc, D. S. (2006). Evidence for clonal evolution among highly polymorphic genes in methicillin-resistant *Staphylococcus aureus*. *Journal of Bacteriology,* 188**,** 169-178.

Lehmann, T., Blackston, C. R., Parmley, S. F., Remington, J. S. & Dubey, J. P. (2000). Strain typing of *Toxoplasma gondii*: comparison of antigen-coding and housekeeping genes. *Journal of Parasitology,* 86**,** 960-971.

Lehmann, T., Graham, D. H., Dahl, E., Sreekumar, C., Launer, F., Corn, J. L., Gamble, H. R. & Dubey, J. P. (2003). Transmission dynamics of *Toxoplasma gondii* on a pig farm. *Infection Genetics and Evolution,* 3**,** 135-141.

Lehmann, T., Graham, D. H., Dahl, E. R., Bahia-Oliveira, L. M. G., Gennari, S. M. & Dubey, J. P. (2004). Variation in the structure of *Toxoplasma gondii* and the roles of selfing, drift, and epistatic selection in maintaining linkage disequilibria. *Infection Genetics and Evolution,* 4**,** 107-114.

Lekutis, C., Ferguson, D. J. P., Grigg, M. E., Camps, M. & Boothroyd, J. C. (2001). Surface antigens of *Toxoplasma gondii*: variations on a theme. *International Journal for Parasitology,* 31**,** 1285-1292.

Lindstroem, I., Sundar, N., Lindh, J., Kironde, F., Kabasa, J. D., Kwok, O. C. H., Dubey, J. P. & Smith, J. E. (2008). Isolation and genotyping of *Toxoplasma gondii* from Ugandan chickens reveals frequent multiple infections. *Parasitology,* 135**,** 39-45.

Luft, B. J. & Remington, J. S. (1992). Toxoplasmic encephalitis in AIDS. *Clinical Infectious Diseases,* 15**,** 211-222.

Maiden, M. C. J. (2006). Multilocus sequence typing of bacteria. *Annual Review of Microbiology,* 60, 561-588.

Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J. J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M. & Spratt, B. G. (1998). Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 3140-3145.

Mallon, M., Macleod, A., Wastling, J., Smith, H., Reilly, B. & Tait, A. (2003). Population structures and the role of genetic exchange in the zoonotic pathogen *Cryptosporidium parvum*. *Journal of Molecular Evolution,* 56**,** 407-417.

Marcello, L. & Barry, J. D. (2007). Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Research,* 17**,** 1344-1352.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z. T., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., Mcdade, K. E., Mckenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P. G., Begley, R. F. & Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-380.

Mccabe, R. & Remington, J. S. (1988). Toxoplasmosis: the time has come. *New England Journal of Medicine,* 318**,** 313-315.

Meisel, R., Stachelhaus, S., Mevelec, M. N., Reichmann, G., Dubremetz, J. F. & Fischer, H. G. (1996). Identification of two alleles in the GRA4 locus of *Toxoplasma gondii* determining a differential epitope which allows discrimination of type I versus type II and III strains. *Molecular and Biochemical Parasitology,* 81**,** 259-263.

Melo, M. B., Jensen, K. D. C. & Saeij, J. P. J. (2011). *Toxoplasma gondii* effectors are master regulators of the inflammatory response. *Trends in Parasitology,* 27**,** 487-495.

Mercier, C., Howe, D. K., Mordue, D., Lingnau, M. & Sibley, L. D. (1998). Targeted disruption of the GRA2 locus in *Toxoplasma gondii* decreases acute virulence in mice. *Infection and Immunity,* 66**,** 4176-4182.

Mercier, C., Rauscher, B., Lecordier, L., Deslée, D., Dubremetz, J.-F. & Cesbron-Delauw, M.-F. (2001). Lack of expression of the dense granule protein GRA5 does not affect the development of Toxoplasma tachyzoites. *Molecular and Biochemical Parasitology,* 116**,** 247-251.

Mets, M. B., Holfels, E., Boyer, K. M., Swisher, C. N., Roizen, N., Stein, L., Stein, M., Hopkins, J., Withers, S., Mack, D., Luciano, R., Patel, D., Remington, J. S., Meier, P. & Mcleod, R. (1996). Eye manifestations of congenital toxoplasmosis. *American Journal of Ophthalmology,* 122**,** 309-324.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics,* 11**,** 31-46.

Miller, M. A., Grigg, M. E., Kreuder, C., James, E. R., Melli, A. C., Crosbie, P. R., Jessup, D. A., Boothroyd, J. C., Brownstein, D. & Conrad, P. A. (2004). An unusual genotype of *Toxoplasma gondii* is common in California sea otters (Enhydra lutris nereis) and is a cause of mortality. *International Journal for Parasitology,* 34**,** 275-284.

Montoya, J. G. & Remington, J. S. (1996). Toxoplasmic chorioretinitis in the setting of acute acquired toxoplasmosis. *Clinical Infectious Diseases,* 23**,** 277-282.

Mordue, D. G., Monroy, F., La Regina, M., Dinarello, C. A. & Sibley, L. D. (2001). Acute toxoplasmosis leads to lethal overproduction of Th1 cytokines. *Journal of Immunology,* 167**,** 4574-4584.

Nagel, S. D. & Boothroyd, J. C. (1989). The major surface antigen, P30, of *Toxoplasma gondii* is anchored by a glycolipid. *Journal of Biological Chemistry,* 264**,** 5569-5574.

Nguyen, T. D., Bigaignon, G., Markine-Goriaynoff, D., Heremans, H., Nguyen, T. N., Warnier, G., Delmee, M., Warny, M., Wolf, S. F., Uyttenhove, C., Van Snick, J. & Coutelier, J. P. (2003). Virulent *Toxoplasma gondii* strain RH promotes T-cell-

independent overproduction of proinflammatory cytokines IL12 and gamma-interferon. *Journal of Medical Microbiology,* 52**,** 869-876.

Nicholson, D. H. & Wolchok, E. B. (1976). Ocular toxoplasmosis in an adult receiving long-term corticosteroid therapy. *Archives of Ophthalmology,* 94**,** 248-254.

Niedelman, W., Gold, D. A., Rosowski, E. E., Sprokholt, J. K., Lim, D., Arenas, A. F., Melo, M. B., Spooner, E., Yaffe, M. B. & Saeij, J. P. (2012). The rhoptry proteins ROP18 and ROP5 mediate *Toxoplasma gondii* evasion of the murine, but not the human, interferon-gamma response. *PLoS Pathogens,* 8**,** e1002784.

Ong, Y. C., Reese, M. L. & Boothroyd, J. C. (2010). Toxoplasma rhoptry protein 16 (ROP16) subverts host function by direct tyrosine phosphorylation of Stat6. *Journal of Biological Chemistry,* 285**,** 28731-28740.

Ottaviani, A., Gilson, E. & Magdinier, F. (2008). Telomeric position effect: from the yeast paradigm to human pathologies? *Biochimie,* 90**,** 93-107.

Parmley, S., Goebel, F. & Remington, J. (1992). Detection of *Toxoplasma gondii* in cerebrospinal fluid from AIDS patients by polymerase chain reaction. *Journal of Clinical Microbiology,* 30**,** 3000-3002.

Peixoto, L., Chen, F., Harb, O. S., Davis, P. H., Beiting, D. P., Brownback, C. S., Ouloguem, D. & Roos, D. S. (2010). Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses. *Cell Host & Microbe,* 8**,** 208-218.

Pena, H. F. J., Gennari, S. M., Dubey, J. P. & Su, C. (2008). Population structure and mouse-virulence of *Toxoplasma gondii* in Brazil. *International Journal for Parasitology,* 38**,** 561-569.

Pernas, L. & Boothroyd, J. C. (2010). Association of host mitochondria with the parasitophorous vacuole during Toxoplasma infection is not dependent on rhoptry proteins ROP2/8. *International Journal for Parasitology,* 40**,** 1367-1371.

Pfefferkorn, E. R., Pfefferkorn, L. C. & Colby, E. D. (1977). Development of gametes and oocysts in cats fed cysts derived from cloned trophozoites of *Toxoplasma gondii*. *Journal of Parasitology,* 63**,** 158-159.

Porter, S. B. & Sande, M. A. (1992). Toxoplasmosis of the central nervous system in the acquired immunodeficiency syndrome. *New England Journal of Medicine,* 327**,** 1643-1648.

Pryde, F. E. & Louis, E. J. (1999). Limitations of silencing at native yeast telomeres. *EMBO Journal,* 18**,** 2538-2550.

Radke, J. R., Striepen, B., Guerini, M. N., Jerome, M. E., Roos, D. S. & White, M. W. (2001). Defining the cell cycle for the tachyzoite stage of *Toxoplasma gondii*. *Molecular and Biochemical Parasitology,* 115**,** 165-175.

Rajendran, C., Su, C. & Dubey, J. P. (2012). Molecular genotyping of *Toxoplasma gondii* from Central and South America revealed high diversity within and between populations. *Infection Genetics and Evolution,* 12**,** 359-368.

Reese, M. L., Shah, N. & Boothroyd, J. C. (2014). The Toxoplasma pseudokinase ROP5 is an allosteric inhibitor of the immunity-related GTPases. *Journal of Biological Chemistry,* 289, 27849-27858.

Reese, M. L., Zeiner, G. M., Saeij, J. P., Boothroyd, J. C. & Boyle, J. P. (2011). Polymorphic family of injected pseudokinases is paramount in Toxoplasma virulence. *Proceedings of the National Academy of Sciences,* 108, 9625-9630.

Roberts, F. & Mcleod, R. (1999). Pathogenesis of toxoplasmic retinochoroiditis. *Parasitology Today,* 15**,** 51-57.

Robinson, D. A., Monk, A. B., Cooper, J. E., Feil, E. J. & Enright, M. C. (2005). Evolutionary genetics of the accessory gene regulator (agr) locus in *Staphylococcus aureus*. *Journal of Bacteriology,* 187**,** 8312-8321.

Robinson, S. A., Smith, J. E. & Millner, P. A. (2004). *Toxoplasma gondii* major surface antigen (SAGI): in vitro analysis of host cell binding. *Parasitology,* 128**,** 391-396.

Romand, S., Chosson, M., Franck, J., Wallon, M., Kieffer, F., Kaiser, K., Dumon, H., Peyron, F., Thulliez, P. & Picot, S. (2004). Usefulness of quantitative polymerase chain reaction in amniotic fluid as early prognostic marker of fetal infection with *Toxoplasma gondii*. *American Journal of Obstetrics and Gynecology,* 190**,** 797-802.

Rome, M. E., Beck, J. R., Turetzky, J. M., Webster, P. & Bradley, P. J. (2008). Intervacuolar transport and unique topology of GRA14, a novel dense granule protein in *Toxoplasma gondii*. *Infection and immunity,* 76**,** 4865-4875.

Rosowski, E. E., Lu, D., Julien, L., Rodda, L., Gaiser, R. A., Jensen, K. D. & Saeij, J. P. (2011). Strain-specific activation of the NF-κB pathway by GRA15, a novel *Toxoplasma gondii* dense granule protein. *The Journal of Experimental Medicine,* 208, 195-212.

Sabin, A. B. (1941). Toxoplasmic encephalitis in children. *Journal of the American Medical Association,* 116**,** 801.

Sadak, A., Taghy, Z., Fortier, B. & Dubremetz, J. F. (1988). Characterization of a family of rhoptry proteins of *Toxoplasma gondii*. *Molecular and Biochemical Parasitology,* 29**,** 203-211.

Saeij, J. P. J., Boyle, J. P. & Boothroyd, J. C. (2005). Differences among the three major strains of *Toxoplasma gondii* and their specific interactions with the infected host. *Trends in Parasitology,* 21**,** 476-481.

Saeij, J. P. J., Boyle, J. P., Coller, S., Taylor, S., Sibley, L. D., Brooke-Powell, E. T., Ajioka, J. W. & Boothroyd, J. C. (2006). Polymorphic secreted kinases are key virulence factors in toxoplasmosis. *Science,* 314**,** 1780-1783.

Sakharkar, K. R., Dhar, P. K. & Chow, V. T. K. (2004). Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. *International Journal of Systematic and Evolutionary Microbiology,* 54**,** 1937-1941.

Sibley, L. D. (2003). Recent origins among ancient parasites. *Veterinary Parasitology,* 115**,** 185-198.

Sibley, L. D. (2011). Invasion and intracellular survival by protozoan parasites. *Immunological Reviews,* 240**,** 72-91.

Sibley, L. D. & Boothroyd, J. C. (1992). Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature,* 359**,** 82-85.

Sibley, L. D., Khan, A., Ajioka, J. W. & Rosenthal, B. M. (2009). Genetic diversity of *Toxoplasma gondii* in animals and humans. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 364**,** 2749-2761.

Sibley, L. D., Mordue, D. & Howe, D. K. (1999). Experimental approaches to understanding virulence in toxoplasmosis. *Immunobiology,* 201**,** 210-224.

Sibley, L. D., Mordue, D. G., Su, C. L., Robben, P. M. & Howe, D. K. (2002). Genetic approaches to studying virulence and pathogenesis in *Toxoplasma gondii*. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 357**,** 81-88.

Sreekumar, C., Graham, D. H., Dahl, E., Lehmann, T., Raman, M., Bhalerao, D. P., Vianna, M. C. B. & Dubey, J. P. (2003). Genotyping of *Toxoplasma gondii* isolates from chickens from India. *Veterinary Parasitology,* 118**,** 187-194.

Steinfeldt, T., Konen-Waisman, S., Tong, L., Pawlowski, N., Lamkemeyer, T., Sibley, L. D., Hunn, J. P. & Howard, J. C. (2010). Phosphorylation of mouse immunity-related

GTPase (IRG) resistance proteins is an evasion strategy for virulent *Toxoplasma gondii*. *PLoS Biology,* 8, e1000576.

Stringer, J. R. & Cushion, M. T. (1998). II. The genome of *Pneumocystis carinii*. *FEMS Immunology and Medical Microbiology,* 22**,** 15-26.

Su, C., Evans, D., Cole, R. H., Kissinger, J. C., Ajioka, J. W. & Sibley, L. D. (2003). Recent expansion of Toxoplasma through enhanced oral transmission. *Science,* 299**,** 414-416.

Su, C., Khan, A., Zhou, P., Majumdar, D., Ajzenberg, D., Darde, M.-L., Zhu, X.-Q., Ajioka, J. W., Rosenthal, B. M., Dubey, J. P. & Sibley, L. D. (2012). Globally diverse *Toxoplasma gondii* isolates comprise six major clades originating from a small number of distinct ancestral lineages. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 5844-5849.

Su, C., Shwab, E. K., Zhou, P., Zhu, X. Q. & Dubey, J. P. (2010). Moving towards an integrated approach to molecular detection and identification of *Toxoplasma gondii*. *Parasitology,* 137**,** 1-11.

Su, C., Zhang, X. & Dubey, J. P. (2006). Genotyping of *Toxoplasma gondii* by multilocus PCR-RFLP markers: A high resolution and simple method for identification of parasites. *International Journal for Parasitology,* 36**,** 841-848.

Su, C. L., Howe, D. K., Dubey, J. P., Ajioka, J. W. & Sibley, L. D. (2002). Identification of quantitative trait loci controlling acute virulence in *Toxoplasma gondii*. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 10753-10758.

Suzuki, Y., Wong, S. Y., Grumet, F. C., Fessel, J., Montoya, J. G., Zolopa, A. R., Portmore, A., Schumacherperdreau, F., Schrappe, M., Koppen, S., Ruf, B., Brown, B. W. & Remington, J. S. (1996). Evidence for genetic regulation of susceptibility to toxoplasmic encephalitis in AIDS patients. *Journal of Infectious Diseases,* 173**,** 265-268.

Taylor, L. H., Welburn, S. C. & Woolhouse, M. E. J. (2002). *Theileria annulata*: virulence and transmission from single and mixed clone infections in cattle. *Experimental Parasitology,* 100**,** 186-195.

Taylor, S., Barragan, A., Su, C., Fux, B., Fentress, S. J., Tang, K., Beatty, W. L., El Hajj, H., Jerome, M., Behnke, M. S., White, M., Wootton, J. C. & Sibley, L. D. (2006). A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii*. *Science,* 314**,** 1776-1780.

Templeton, A. R., Crandall, K. A. & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics,* 132**,** 619-633.

Tibayrenc, M. (2009). Multilocus enzyme electrophoresis for parasites and other pathogens. *Molecular Epidemiology of Microorganisms.* Springer.

Tibayrenc, M. & Ayala, F. J. (2002). The clonal theory of parasitic protozoa: 12 years on. *Trends in Parasitology,* 18**,** 405-410.

Tibayrenc, M., Kjellberg, F., Arnaud, J., Oury, B., Breniere, S. F., Darde, M. L. & Ayala, F. J. (1991). Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. *Proceedings of the National Academy of Sciences of the United States of America,* 88**,** 5129-5133.

Tibayrenc, M., Kjellberg, F. & Ayala, F. J. (1990). A clonal theory of parasitic protozoa: the population structures of Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas, and Trypanosoma and their medical and taxonomical consequences. *Proceedings of the National Academy of Sciences of the United States of America,* 87**,** 2414-2418.

Tibayrenc, M., Neubauer, K., Barnabe, C., Guerrini, F., Skarecky, D. & Ayala, F. J. (1993). Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 1335-1339.

Vallochi, A. L., Muccioli, C., Martins, M. C., Silveira, C., Belfort, R. & Rizzo, L. V. (2005). The genotype of *Toxoplasma gondii* strains causing ocular toxoplasmosis in humans in Brazil. *American Journal of Ophthalmology,* 139**,** 350-351.

Villena, I., Marle, M., Darde, M. L., Pinon, J. M. & Aubert, D. (2004). Toxoplasma strain type and human disease: risk of bias during parasite isolation? *Trends in Parasitology,* 20**,** 160-162

Webster, J. P. (2001). Rats, cats, people and parasites: the impact of latent toxoplasmosis on behaviour. *Microbes and Infection,* 3**,** 1037-1045.

Wendte, J. M., Gibson, A. K. & Grigg, M. E. (2011). Population genetics of *Toxoplasma gondii*: New perspectives from parasite genotypes in wildlife. *Veterinary Parasitology,* 182**,** 96-111.

Wendte, J. M., Miller, M. A., Lambourn, D. M., Magargal, S. L., Jessup, D. A. & Grigg, M. E. (2010). Self-mating in the definitive host potentiates clonal outbreaks of the apicomplexan parasites *Sarcocystis neurona* and *Toxoplasma gondii*. *PLoS Genetics,* 6, e1001261.

Wong, S. Y. & Remington, J. S. (1994). Toxoplasmosis in pregnancy. *Clinical Infectious Diseases,* 18**,** 853-861.

Yamamoto, M., Standley, D. M., Takashima, S., Saiga, H., Okuyama, M., Kayama, H., Kubo, E., Ito, H., Takaura, M., Matsuda, T., Soldati-Favre, D. & Takeda, K. (2009). A single polymorphic amino acid on *Toxoplasma gondii* kinase ROP16 determines the direct and strain-specific activation of Stat3. *Journal of Experimental Medicine,* 206**,** 2747-2760.

Zenner, L., Foulet, A., Caudrelier, Y., Darcy, F., Gosselin, B., Capron, A. & Cesbron-Delauw, M. F. (1999). Infection with *Toxoplasma gondii* RH and Prugniaud strains in mice, rats and nude rats: Kinetics of infection in blood and tissues related to pathology in acute and chronic infection. *Pathology Research and Practice,* 195**,** 475-485.

APPENDICES

Appendix 1

A customised set of commands that were used to call SNPs in TgCkUg8 and 9 mapped reads against the Type II reference strain (TgME49) with 3 and 5 reads depths for each sample. (This was performed by Dr. Rachel Brenchley at the University of Salford, using BCFtools (http://samtools.github.io/bcftools/))

Commands:

- samtools faidx refGenome_toxogondii.fa
- samtools mpileup -g -f blast_databases/toxoplasma/refGenome_toxogondii.fa SOL6058.sorted.bam >toxo6058_raw.bcf
- samtools mpileup -g -f blast_databases/toxoplasma/refGenome_toxogondii.fa SOL6059.sorted.bam >toxo6059_raw.bcf
- bcftools view -bvcg toxo6058_raw.bcf >toxo6058_var.bcf
- bcftools view -bvcg toxo6059_raw.bcf >toxo6059_var.bcf

Filtration of SNPs with minimum depth of 3 reads:

Commands:

TgCkUg8:
bcftools view toxo6058_var.bcf | vcfutils.pl varFilter - >toxo6058.vcf
grep -v '^#' toxo6058.vcf | wc -l

TgCkUg9:
bcftools view toxo6059_var.bcf | vcfutils.pl varFilter - >toxo6059.vcf
grep -v '^#' toxo6059.vcf | wc -l

Filtration of SNPs with minimum depth of 5 reads:

Commands:

TgCkUg8:
bcftools view toxo6058_var.bcf | vcfutils.pl varFilter -d 5 | grep -v '^#' >toxo6058_filtered5x.vcf
wc -l toxo6058_filtered5x.vcf

TgCkUg9:
bcftools view toxo6059_var.bcf | vcfutils.pl varFilter -d 5 | grep -v '^#' >toxo6059_filtered5x.vcf
wc -l toxo6059_filtered5x.vcf

To obtain a list of SNPs in each chromosome to compare between the samples:

Commands:

```
awk '{ print $1’_’$2 }' toxo6058_filtered5x.vcf >toxo6058_chrAndPos.txt

awk '{ print $1’_’$2 }' toxo6059_filtered5x.vcf >toxo6059_chrAndPos.txt

cat toxo6058_chrAndPos.txt toxo6059_chrAndPos.txt | sort | uniq -d | wc -l
```

Appendix 2

A customised set of commands that were used to call SNPs within the genes of SRS, ROPs and GRA families in TgCkUg8 and 9 mapped reads against the type II reference strain genome sequences of *Toxoplasma gondii* (TgME49) in fasta format (retrieved from TOXODB)
(http://toxodb.org/common/downloads/Current_Release/TgondiiME49/fasta/data/)
and annotated genome file of TgME49 in gff format (retrieved from TOXODB) (http://www.toxodb.org/common/downloads/Current_Release/TgondiiME49/gff/data/)
with 3 and 5 reads depths for each sample. (This was performed by Dr. Rachel Brenchley at the University of Salford, using BCFtools
(http://samtools.github.io/bcftools/)).

Filtration of SNPs within SRS genes in TgCkUg8 & 9 with minimum depth of 3 reads:
Commands:
TgCkUg8:
./toxo_snps_in_genes.py toxo_SRS_genes.gff toxo6058.vcf

TgCkUg9:
./toxo_snps_in_genes.py toxo_SRS_genes.gff toxo6059.vcf

Filtration of SNPs within SRS genes in TgCkUg8 & 9 with minimum depth of 5 reads:
Commands:
TgCkUg8:
./toxo_snps_in_genes.py toxo_SRS_genes.gff toxo6058_filtered5x.vcf

TgCkUg9:
./toxo_snps_in_genes.py toxo_SRS_genes.gff toxo6059_filtered5x.vcf

Filtration of SNPs within ROP genes in TgCkUg8 & 9 with minimum depth of 3 reads:
Commands:
TgCkUg8:
./toxo_snps_in_genes.py toxo_ROP_genes.gff toxo6058.vcf

TgCkUg9:
./toxo_snps_in_genes.py toxo_ROP_genes.gff toxo6059.vcf

Filtration of SNPs within ROP genes in TgCkUg8 & 9 with minimum depth of 5 reads:

Commands:

TgCkUg8:

./toxo_snps_in_genes.py toxo_ROP_genes.gff toxo6058_filtered5x.vcf

TgCkUg9:

./toxo_snps_in_genes.py toxo_ROP_genes.gff toxo6059_filtered5x.vcf

Filtration of SNPs within GRA genes in TgCkUg8 & 9 with minimum depth of 3 reads:

Commands:

TgCkUg8:

./toxo_snps_in_genes.py toxo_GRA_genes.gff toxo6058.vcf

TgCkUg9:

./toxo_snps_in_genes.py toxo_GRA_genes.gff toxo6059.vcf

Filtration of SNPs within GRA genes in TgCkUg8 & 9 with minimum depth of 5 reads:

Commands:

TgCkUg8:

./toxo_snps_in_genes.py toxo_GRA_genes.gff toxo6058_filtered5x.vcf

TgCkUg9:

./toxo_snps_in_genes.py toxo_GRA_genes.gff toxo6059_filtered5x.vcf

Appendix 3

Comparison through alignment between the MiSeq sequences for TgCkUg8 and 9 strains with the Sanger sequence data for specific loci generated in chapter 2, and the Sanger sequence data for 34 loci generated by Bontell et al., (2009).

TgCkUg8:

```
TgCkUg8 SAG1 TGTAGTCACTGCTGATTCTCGCTGTTCTCGGCAAGGGCTGACGACCGGAGTACAGTTTTTGTGGGCAGAGCCGCTGTGCA 80
TgCkUg8      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 80

TgCkUg8 SAG1 GCTTTCCGTTGTTCTCGGTTGTGTCACATGTGTCATTGTCGTGTAAACACACGGTTGTATGTCGGTTTCGCTGCACCACT 160
TgCkUg8      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 160

TgCkUg8 SAG1 TCATTATTTCTTCTGGTTTTTTGGCGAGTATGTTTCCGAAGGCAGTGAGACGCGCCGTCACGGCAGGGGTGTTTGCCGCG 240
TgCkUg8      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGTTTCCGAAGGCAGTGAGACGCGCCGTCACGGCAGGGGTGTTTGCCGCG 240

TgCkUg8 SAG1 CCCACACTGATGTCGTTCTTGCGATGTGGCGCTATGGCATCGGATCCCCCTCTTGTTGCCAATCAAGTTGTCACCTGCCC 320
TgCkUg8      CCCACACTGATGTCGTTCTTGCGATGTGGCGCTATGGCATCGGATCCCCCTCTTGTTGCCAATCAAGTTGTCACCTGCCC 320

TgCkUg8 SAG1 AGATAAAAAATCGACAGCCGCGGTCATT 348
TgCkUg8      AGATAAAAAATCGACAGCCGCGGTCATT 348
```

```
TgCkUg8 SAG2 AGCGCTGCTTGCGATTCTGTGTGTTGTTTCACGGGTTGCAGTTCTAGGAACTGAGTTGTGATTGTGCACAATTGCGGTGT 80
TgCkUg8      AGCGCTGCTTGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 80

TgCkUg8 SAG2 GACACCTTCTGTCTCGTTCCAATCTTTGTCTTGTCGGAACTATGAGTTTCTCAAAGACCACGAGCCTAGCGTCGCTAGCG 160
TgCkUg8      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 160

TgCkUg8 SAG2 CTCACGGGCTTGTTTGTTGTGTTCAAGTTCGCTCTTGCGTCCACCACCGAGACGCCAGCGCCCATTGAGTGCACTGCCGG 240
TgCkUg8      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTGCCGG 240

TgCkUg8 SAG2 CGCAACGAAGACTGTTGAGGCACCCTCCAGTGGTTCCGTTGTCTTCCAATGTGGGGATAAACTAACCATCAGTCCCAGTG 320
TgCkUg8      CGCAACGAAGACTGTTGAGGCACCCTCCAGTGGTTCCGTTGTCTTCCAATGTGGGGATAAACTAACCATCAGTCCCAGTG 320

TgCkUg8 SAG2 GCGAAGGTGATGTCTTTTATGGCAAGGAATGCACAGACTCGAGGAAGTTGACGACTGTCCTTCCAGGTGCGGTCTTGAAA 400
TgCkUg8      GCGAAGGTGATGTCTTTTATGGCAAGGAATGCACAGACTCGAGGAAGTTGACGACTGTCCTTCCAGGTGCGGTCTTGAAA 400

TgCkUg8 SAG2 GCTAAGGTCGAGCAGCCCCCGAAAGGTCCTGCTACCTACACACTGTCTTACGACGGTACTCCCGAGAAACCTCAGGTTCT 480
TgCkUg8      GCTAAGGTCGAGCAGCCCCCGAAAGGTCCTGCTACCTACACACTGTCTTACGACGGTACTCCCGAGAAACCTCAGGTTCT 480

TgCkUg8 SAG2 CTGTTACAAGTGCGTTGCCGAAGCAG 506
TgCkUg8      CTGTTACAAGTGCGTTGCCGAAGCAG 506
```

```
TgCkUg8 BTUB CGGCATTGATCCGGTGAGGGAGAACGGGTTCCGGAGTGCGGGTTCCCGGACGCTGTCTTTTGTCCTGTATTCTGCGTGAC 80
TgCkUg8      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 80

TgCkUg8 BTUB AGCTTCGCCAGTGTAAATGGAGCCGTTTTCCCTGTGGAGAATCGCGGAGAATGGAGGACTTTTCCGAGCTCGCAGGTGCC 160
TgCkUg8      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGGAGAATCGCGGAGAATGGAGGACTTTTCCGAGCTCGCAGGTGCC 160

TgCkUg8 BTUB ACCCTCCACCTCGACGCAACCAAGTGCACGCACATTTGCCGGGTGGTGACCCTAAGACACCGCAGGTCTACCTGCGCGTT 240
TgCkUg8      ACCCTCCACCTCGACGCAACCAAGTGCACGCACATTTGCCGGGTGGTGACCCTAAGACACCGCAGGTCTACCTGCGCGTT 240

TgCkUg8 BTUB TTTTTCAGTCTTTGCACACAGTTGCACCGAAAGTCATGTTTTTTGCGAAGACATCGGTCGTTCTGGTGGGGGAATACTTC 320
TgCkUg8      TTTTTCAGTCTTTGCACACAGTTGCACCGAAAGTCATGTTTTTTGCGAAGACATCGGTCGTTCTGGTGGGGGAATACTTC 320

TgCkUg8 BTUB ACTCCTGTTCGCGCCTATGTGCGCAGACAGGTGTCCACCCTCCGCATACGG 371
TgCkUg8      ACTCCTGTTCGCGCCTATGTGCGCAGACAGGTGTCCACCCTCCGCATACGG 371
```

```
TgCkUg8 GRA6 GGGTCGCTTTTTTGAAACAGCAGGAAAACAGCTTCGTGGTGCCACGTAGCGTGCTTGTTGGCGACTACCTTTTTTTCTTG 80
TgCkUg8      GGGTCGCTTTTTTGAAACAGCAGGAAAACAGCTTCGTGGTGCCACGTAGCGTGCTTGTTGGCGACTACCTTTTTTTCTTG 80

TgCkUg8 GRA6 GGAGTGTCGGCGAAATGGCACACGGTGGCATCTATCTGAGGCAGAAGCGTAACTTCTGTCCTTTAACTGTCTCCACAGTT 160
TgCkUg8      GGAGTGTCGGCGAAATGGCACACGGTGGCATCTATCTGAGGCAGAAGCGTAACTTCTGTCCTTTAACTGTCTCCACAGTT 160

TgCkUg8 GRA6 GCTGTGGTCTTTGTAGTCTTCATGGGTGTACTCGTCAATTCGTTGGGTGGAGTCGCTGTCGCAGCAGACAGCGGTGGTGT 240
TgCkUg8      GCTGTGGTCTTTGTAGTCTTCATGGGTGTACTCGTCAATTCGTTGGGTGGAGTCGCTGTCGCAGCAGACAGCGGTGGTGT 240
```

```
TgCkUg8 GRA6 TAGGCAGACCCCTTCGGAAACCGGTTCGAGCGGTGGACAGCAAGAAGCAGTGGGGACCACTGAAGA 306
TgCkUg8      TAGGCAGACCCCTTCGGAAACCGGTTCGAGCGGTGGACAGCAAGAAGCAGTGGGGACCACTGAAGA 306

TgCkUg8 C22-8 TTTCTCGTTCCTTTCGCGCTGCGTTTCGTTTCTTTTTCAAAAACGGACGGCATGCGGTTCGCATTCCCGCCGTGAAGCAA 80
TgCkUg8       TTTCTCGTTCCTTTCGCGCTGCGTTTCGTTTCTTTTTCANNNNNNNNNNNNNNNNNNNNNNNNNCCCGCCGTGAAGCAA 80

TgCkUg8 C22-8 TCGAAGAATTTCCGAAGTCAAGAAGAGCGAACCTTCTGTCATCTCCCCTGCCTTTCTCTCAGATCCCTTCGCTTCTTTGT 160
TgCkUg8       TCGAAGAATTTCCGAAGTCAAGAAGAGCGAACCTTCTGTCATCTCCCCTGCCTTTCTCTCAGATCCCTTCGCTTCTTTGT 160

TgCkUg8 C22-8 GCAGGGTTTTGGAGAGATTGAAATTCCTCGAGAAGCAAATTCGGAGGTGCCACTGGAAGGCAACCGCTGCCTCCCACAAC 240
TgCkUg8       GCAGGGTTTTGGAGAGATTGAAATTCCTCGAGAAGCAAATTCGGAGGTGCCACTGGAAGGCAACCGCTGCCTCCCACAAC 240

TgCkUg8 C22-8 ACTGGAAACGCACTGGAACTTTTCAGCATTTTCGCTGAACGGAAAATGCAGCAAACCTGTGAACTCGCACAGCTGTAGCG 320
TgCkUg8       ACTGGAAACGCACTGGAACTTTTCAGCATTTTCGCTGAACGGAAAATGCAGCAAACCTGTGAACTCGCACAGCTGTAGCG 320

TgCkUg8 C22-8 ACGAGCACTACCCCGAAAAACGACCCTCGAGACTGCATGCATGCAGTTCATGAGGAACCGGAAAGGAAAAAGACACTTTC 400
TgCkUg8       ACGAGCACTACCCCGAAAAACGACCCTCGAGACTGCATGCATGCAGTTCATGAGGAACCGGAAAGGAAAAAGACACTTTC 400

TgCkUg8 C22-8 CCCGATCCTTCTCTACGCATCAAAGTCGAATTGAATCTAAAAACGAGAGAAATCGATTTGCTTCAAACTCTGAGTTCCAA 480
TgCkUg8       CCCGATCCTTCTCTACGCATCAAAGTCGAATTGAATCTAAAAACGAGAGAAATCGATTTGCTTCAAACTCTGAGTTCCAA 480

TgCkUg8 C22-8 GTTCT 485
TgCkUg8       GTTCT 485

TgCkUg8 C29-2 TTTTTCCGGAAGCGGAGCATGTGGACACGTATGCCATTTCAAATGAATCGAGGACCGGAGGGCTAACGAGACGCCAGTGT 80
TgCkUg8       NTTTTCCGGAAGCGGAGCATGTGGACACGTATGCCATTTCAAATGAATCGAGGACCGGAGGGCTAACGAGACGCCAGTGT 80

TgCkUg8 C29-2 ACACACTGGAGGAGGTCCGTCCATTCCTGTTCCTCGTTTTCCGATTCTGTGTGTCTTTCCCATTTTCGCGCTCTTCCACT 160
TgCkUg8       ACACACTGGAGGAGGTCCGTCCATTCCTGTTCCTCGTTTTCCGATTCTGTGTGTCTTTCCCATTTTCGCGCTCTTCCACT 160

TgCkUg8 C29-2 TCTATGCGTTTATTTGATTGCCTTCCAGCCATGCTGTTCAACAAGAGTGGCTTTTCGCGGGAGTAAGAACTCAGCCACAT 240
TgCkUg8       TCTATGCGTTTATTTGATTGCCTTCCAGCCATGCTGTTCAACAAGAGTGGCTTTTCGCGGGAGTAAGAACTCAGCCACAT 240

TgCkUg8 C29-2 GCAGTACACAAAAGTCCCTCGAATGCTCTCTGAACACACCGAACTCCCGGGAAGGTGCTCACAAACATGCCCACCACCGG 320
TgCkUg8       GCAGTACACAAAAGTCCCTCGAATGCTCTCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 320

TgCkUg8 C29-2 ATATCCCGTTGTGCTAAAGTGTTTCATCAACTTCTCTCTCTTGATATCCATTCTCACATCACTCCTACGTATTCTGCTCA 400
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNTCATCAACTTCTCTCTCTTGATATCCATTCTCACATCACTCCTACGTATTCTGCTCA 400

TgCkUg8 C29-2 GTCCGAACAG 410
TgCkUg8       GTCCGAACAG 410

TgCkUg8 L358 CGCAGCATGCCGCGAAGAATGGAATGTTCTCTTCCTCGACACGGGCGGAGAAAGCGAAACCTTCCTGCTGCGCGGTTTCC 80
TgCkUg8      CGCAGCATGCCGCGAAGAATGGAATGTTCTCTTCCTCGACACGGGCGGAGAAAGCGAAACCTTCCTGCTGCGCGGTTTCC 80

TgCkUg8 L358 GTCGCTGTCTCGTCGCTCGCTTCTCTCCCCGCTTCTCCGTCCGAAACGAGTCGCGAGAGAGGCGGCAGGCGAATGCAGGA 160
TgCkUg8      GTCGCTGTCTCGTCGCTCGCTTCTCTCCCCGCTTCTCCGTCCGAAACGAGTCGCGAGAGAGGCGGCAGGCGAATGCAGGA 160

TgCkUg8 L358 GAAACTGATCTCTGCCTCTCTCCTCGCCTTCTTGTCTATCAGCGGGGTGGCCGCGCGACATTGCAGCGTCGACGCGCAGC 240
TgCkUg8      GAAACTGATCTCTGCCTCTCTCCTCGCCTTCTTGTCTATCAGCGGGGTGGCCGCGCGACATTGCAGCGTCGACGCGCAGC 240

TgCkUg8 L358 CGAGCGAAGGAAGCGGGGAAGACGAAGAGGCTCCGTTCCCCCGACAGGACGCATCCGACAAGCCGAAGGCAGAAAGAGGA 320
TgCkUg8      CGAGCGAAGGAAGCGGGGAAGACGAAGAGGCTCCGTTCCCCCGACAGGACGCATCCGACAAGCCGAAGGCAGAAAGAGGA 320

TgCkUg8 L358 CATGCATGCGAGTGTGTCGTCCGGAAGAGCAGAGACTCGCACCCTTGCAGGGCCAAGACGC 381
TgCkUg8      CATGCATGCGAGTGTGTCGTCCGGAAGAGCAGAGACTCGCACCCTTGCAGGGCCAAGACGC 381

TgCkUg8 PK1  TCTTCTTCATGGAGTAAAAAACTGTTTGGAATGTGTCGGCAGCGACTCTAATCATCTGCTGCCAAATGGCGAATAACGAG 80
TgCkUg8      NNNNNNNNNNNNNNNNNNAAACTGTTTGGAATGTGTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 80

TgCkUg8 PK1  TGCATGTGACGTATGTTCATCAGTACGTGAAGAACCGTTCTCCAGAACAAGCTGTCTCGTAGAGTGGTCAGTACGTTACC 160
TgCkUg8      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 160

TgCkUg8 PK1  TGGATCTCGACGAAAACGCGGTCAGCATCGGTGATCTCGTTGAATGATCTTTTGTTTATAAGCTTGATTGCAACGATCTC 240
TgCkUg8      NNNNNNTCGACGAAAACGCGGTCAGCATCGGTGATCTCGTTGAATGATCTTTTGTTTATAAGCTTGATTGCAACGATCTC 240

TgCkUg8 PK1  TCCTGGTGGTGACAAAGTGCACGCATTTCGCCACTTTGTTGTCGATGACACAAAAAGAGAGAGTTTTAAGGAAACCCATG 320
TgCkUg8      TCCTGGTGGTGACAAAGTGCACGCATTTCGCCACTTTGTTGTCGATGACACAAAAAGAGAGAGTTTTAAGGAAACCCATG 320

TgCkUg8 PK1  GTTTTGCATCACGGCGAAAACGTGTGCCCAACGAACGCGAAACTGGTACCAGGCCACCAAACATTCGATCCCTGTATTCC 400
TgCkUg8      GTTTTGCATCACGGCGAAAACGTGTGCCCAACGAACGCGAAACTGGTACCAGGCCACCAAACATTCGATCCCTGTATTCC 400

TgCkUg8 PK1  GTACAGATTGCACCTTAACGCAGTTCAATGTGGATCGCCTTATACAAGTTGGCTATAGATCTCAACTAGCAAAACTCTCT 480
TgCkUg8      GTACAGATTGCACCTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAAACTCTCT 480
```

```
TgCkUg8 PK1  CTTTCAGACCTAATACAGGTGTACTTTGGTTTACACTAGAGAAAATTGTGCATGACCTGCACGTGGCAAATTCAGGTGCG 560
TgCkUg8      CTTTCAGACCTAATACAGGTGTACTTTGGTTTACACTAGAGAAAATTGTGCATGACCTGCACGTGGCAAATTCAGGTGCG 560

TgCkUg8 PK1  GAAACAGTCTGGACAAAATCGTCTTCCATTGTGCCCTACTGCTTATGAAAAGCATATGGGCAACAAAAAAGCCCCTGATA 640
TgCkUg8      GAAACAGTCTGGACAAAATCGTCTTCCATTGTGCCCTACTGCTTATGAAAAGCATATGGGCAACAAAAAAGCCCCTGATA 640

TgCkUg8 PK1  CTCGGGTCAGCATGTGACACTGTTGGGATTATGTTGCGCAGTTAATTATTCAGCGAAATGCGGGAAGAACGACACCCATA 720
TgCkUg8      CTCGGGTCAGCATGTGACACTGTTGGGATTATGTTGCGCAGTTAATTATTCAGCGAAATGCGGGAAGAACGACACCCATA 720

TgCkUg8 PK1  CAAATTCTGCATCTTAACTGTCCTACACATAAGGTATTGCGCTCCGAACACGTTGCCACTCGCATCGCAGTAACACTCTC 800
TgCkUg8      CAAATTCTGCATCTTAACTGTCCTACACATAAGGTATTGCGCTCCGAACACGTTGCCACTCGCATCGCAGTAACACTCTC 800

TgCkUg8 PK1  GCAGGGAAACCCACTAATGCATCCTATTCGAGACGC 836
TgCkUg8      GCAGGGAAACCCACTAATGCATCCTATTCGAGACGC 836
```

```
TgCkUg8-XI-UPRT1  TTGTCGATCCCCGATATTCGACAAACGACCAGGAAGAAAGCATTCTCCAGGACATCATCA 60
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60

TgCkUg8-XI-UPRT1  CGAGGTAATCCTTCAACCGAAGTTTGCTTTCCGTGACTCTGCCTGTTGGTTATACTGCGT 120
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120

TgCkUg8-XI-UPRT1  GGCCTTCCCGTCCTGCGGCCCCCTTTCCTCCGCTTGCTGTTTAAATGCTCGTCCTCGTTT 180
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180

TgCkUg8-XI-UPRT1  TCCTTCCTGCCGCATCCCCGTATATTTTAAGGAGAGGGAAACAGGCGTGAGTTGGACGGC 240
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240

TgCkUg8-XI-UPRT1  ATGAAAGTTCTCGGCCTGTATGCCGGTTGTGGCGGTCGTTTGCAGATTGCTTTTTTCTTC 300
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300

TgCkUg8-XI-UPRT1  GAATCGGTGCTGTAACCCTCGCGAAGAACGACGCTGCAAACGACTTCTCGAACTCTCAGT 360
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360

TgCkUg8-XI-UPRT1  CGTGTACTTTACGTGCTTCCTTTCAGGGACCTCCCCCCGCGTTACTCATTTGTATTCACA 420
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCCGCGTTACTCATTTGTATTCACA 420

TgCkUg8-XI-UPRT1  GCTACGAAGTGTCTTGCAAGGTGGATTTCTGCCAGGCTCCATGTCTCACTCGTTGCGTTT 480
TgCkUg8           GCTACGAAGTGTCTTGCAAGGTGGATTTCTGCCAGGCTCCATGTCTCACTCGTTGCGTTT 480

TgCkUg8-XI-UPRT1  TCGGAAAAGTTCATTGTGAACGTTCCCCTTGCGTGTCATGACTTTATCAGGTTTCCCAAT 540
TgCkUg8           TCGGAAAAGTTCATTGTGAACGTTCCCCTTGCGTGTCATGACTTTATCAGGTTTCCCAAT 540

TgCkUg8-XI-UPRT1  GTGGTGCTCATGAAGCAGACGGCTCAGCTTCGAGCGATGATGACCATCATTCGTGGTGAG 600
TgCkUg8           GTGGTGCTCATGAAGCAGACGGCTCAGCTTCGAGCGATGATGACCATCATTCGTGGTGAG 600

TgCkUg8-XI-UPRT1  TGGTGACAGGGTAGAACAGACAGACGAGGAAGGAGAAAAAAAAGGAATCAAGCCACTAAGC 660
TgCkUg8           TGGTGACAGGGTAGAACAGACAGACGAGGAAGGAGAAAAAAAAGGAATCAAGCCACTAAGC 660

TgCkUg8-XI-UPRT1  GGCGACACACTGTTGGTGCCTGAAGCCT 688
TgCkUg8           GGCGACACACTGTTGGTGCCTGAAGCCT 688
```

```
TgCkUg8-II-W35  CGTCGCCCGTCCTGCTGCGGTCTTTTCTCTTCTTCGTGGCTTTCTCTCTCTCTCTCTGTG 60
TgCkUg8         CGTCGCCCGTCCTGCTGCGGTCTTTTCTCTTCTTCGTGGCTTTCTCTCTCTCTCTCTGTG 60

TgCkUg8-II-W35  TCGCTGTGCGGATAAAACAGCGCGCGTCTTTGTGGGAAAGGAACGGCATGTTCCCTTCGT 120
TgCkUg8         TCGCTGTGCGGATAAAACAGCGCGCGTCTTTGTGGGAAAGGAACGGCATGTTCCCTTCGT 120

TgCkUg8-II-W35  TTATGTTCATCTGCATTCTCTTCGCATTTGCGAACTCGAGGTTTTCCAGATATCTCGGAG 180
TgCkUg8         TTATGTTCATCTGCATTCTCTTCGCATTTGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180

TgCkUg8-II-W35  TCCGCTGTCGGAAACA 196
TgCkUg8         NNNNNNNNNNNNNNNN 196
```

```
TgCkUg8-XII-B17  CCAGTTTCGCGAACAAGTGAACAGACACCCGATGCCTACAAAGAGAGGAACGTGAGATGC 60
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60

TgCkUg8-XII-B17  CGGAGTGCTGGATGAAAGACAGTGGCTCACTGGGAAGACAGACACCGACAAGTCCACGCG 120
TgCkUg8          NNNNNNNNNNNNNNNAAAGACAGTGGCTCACTGGGNNNNNNNNNNNNNNNNNNNNNNN 120

TgCkUg8-XII-B17  AACTGTTCAGAAGACGAGAGCTTTGCCGCCTTCCTCTTCCACTTTTCTTCCTCTCTTTGC 180
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180
```

```
TgCkUg8-XII-B17  CCCTTAGTGGCCTCGACTCTTGCTGTACTCCGCTTCCCTGTGTGTCACCGTTCCTTCCTT 240
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240

TgCkUg8-XII-B17  CTTCTATCTTCCCTTCTCTCTCTCTCTCAGAGAACTCCCTGCGTTCCATCCCCTCTCTCC 300
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300

TgCkUg8-XII-B17  TTCTTCAGTTGTTCTCTTGCTGTTGTGTCTCTCCTCCTCTAC 342
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 342


TgCkUg8-XII-32   CTTCCAAGCAAACATGAATACCGTATCCAGTGTTAAGTTTAAACGTAATGCAGCTTAATC 60
TgCkUg8          CTTCCAAGCAAACATGAATACCGTATCCAGTGTTAAGTTTAAACGTAATGCAGCTTAATC 60

TgCkUg8-XII-32   TTCTATGTCGTTAAGCTAATCAAATAAGTTTCGTCGTTATTGATATTTCATTGACATGTT 120
TgCkUg8          TTCTATGTCGTTAAGCTAATCAAATAAGTTTCGTCGTTATTGATATTTCATTGACATGTT 120

TgCkUg8-XII-32   GATGACATAAATACAAACAACCGGTTTTGGATGGAATTACTTATAACACCGCACAATATG 180
TgCkUg8          GATGACATAAATACAAACAACCGGTTTTGGATGGAATTACTTATAACACCGCACAATATG 180

TgCkUg8-XII-32   CTAGAAAATACCATTCAGGTACGATATGAAGTGGTGTTACGAACCGATTGACTGGTATTG 240
TgCkUg8          CTAGAAAATACCATTCAGGTACGATATGAAGTGGTGTTACGAACCGATTGACTGGTATTG 240

TgCkUg8-XII-32   AGTTATCTGGATGTGATAGTTCCATCAAACCAAAAGCCGCTTGTTCTCTCGCTGTTAATA 300
TgCkUg8          AGTTATCTGGATGTGATAGTTCCATCAAACCAAAAGCCGCTTGTTCTCTCGCTGTTAATA 300

TgCkUg8-XII-32   CGAGTGATAACAGTAATCCATATATTACGCCTAGAACATAACCGATGTGGAATAACCTTA 360
TgCkUg8          CGAGTGATAACAGTAATCCATATATTACGCCTAGAACATAACCGATGTGGAATAACCTTA 360

TgCkUg8-XII-32   ACATGTATGTTGCTTGTGGAATAGCACTACCTATAATAATCAAGATCATACTATATACCC 420
TgCkUg8          ACATGTATGTTGCTTGTGGAATAGCACTACCTATAATAATCAAGATCATACTATATACCC 420

TgCkUg8-XII-32   AAATACAGGTGTCTAGCATCAGTTGTTAGC 450
TgCkUg8          AAATACAGGTGTCTAGCATCAGTTGTTAGC 450


TgCkUg8-XII-31   TTGCAGAAGAACAGACGCCCAGCAAGAGGAAGCCACAAGACCTTTTTCGGTGAAGTCCCC 60
TgCkUg8          TTGCAGAAGAACAGACGCCCAGCAAGAGGAAGCCACAAGACCTTTTTCGGTGAAGTCCCC 60

TgCkUg8-XII-31   GCATCGAGGGAACCGCGAACCGAATCGAAGCTGCGTCTTCATTTAGCTGCATGCACGCCT 120
TgCkUg8          GCATCGAGGGAACCGCGAACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120

TgCkUg8-XII-31   CCACCTGGGTACCTCCACAGACGGACTTGTAGAAGGATGCGCGATTCAAACGGGCATGCA 180
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180

TgCkUg8-XII-31   CTCCCAGGCAGAGACTGTCTCCTCCGCAGAAGGGCAGAGACCGCTCGCGCAAGAGAGCAG 240
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240

TgCkUg8-XII-31   ACTGCTTTTCAGAGACTCCCGCAGAGGCCTGCTCCGCACTGGTGGACCTGGTGGCGGCTG 300
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300

TgCkUg8-XII-31   CAGGGCGTAGAACTGTTTGCCTGTTGCTGCATGCGACTTTTCAGGAGTCCGGCGCGGTTT 360
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360

TgCkUg8-XII-31   GGGTGGACGTTTCGGGAGGCGCTCGCGCACTGAAAACGCGCGAGGAAAACATTATCGCGG 420
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420

TgCkUg8-XII-31   AGCAGGACGAACGACTCTGTTTCCTCGCTGGGACCGTGAG 460
TgCkUg8          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 460


TgCkUg8-XI-30    AGCCAGCGGTATTAGGGAAATAGCGTCAGGAGATGAAGGGTGCCGGTTCTCGACCAAACC 60
TgCkUg8          AGCCAGCGGTATTAGGGAAATAGCGTCAGGAGATGAAGGGTGCCGGTTCTCGACCAAACC 60

TgCkUg8-XI-30    GTTCGGTCCACCGGCATCCTGCTGCCGGTGCAAGATGCGCAAGACTATATTTTTGAGTTG 120
TgCkUg8          GTTCGGTCCACCGGCATCCTGCTGCCGGTGCAAGATGCGCAAGACTATATTTTTGAGTTG 120

TgCkUg8-XI-30    CTGTAGATCAGAAAAGCGCACAAAAATCATATGAACACCGAATTACCTTTGGCGGAGGGG 180
TgCkUg8          CTGTAGATCAGAAAAGCGCACAAAAATCATATGAACACCGAATTACCTTTGGCGGAGGGG 180

TgCkUg8-XI-30    CTCCTCAAAGTATGGGCTGGTTTGCATTGGATTTGCTTGATGCAATAGCCTTATGTACCC 240
TgCkUg8          CTCCTCAAAGTATGGGCTGGTTTGCATTGGATTTGCTTGATGCAATAGCCTTATGTACCC 240

TgCkUg8-XI-30    AAAACCTCTTCGAGCCTTGTACACATCGCCCGTTACGCCACGGAAATTGATTATTTTAAA 300
TgCkUg8          AAAACCTCTTCGAGCCTTGTACACATCGCCCGTTACGCCACGGAAATTGATTATTTTAAA 300
```

```
TgCkUg8-XI-30  AAACAGTTTCACCCTTGCGAGCGTACTACTCAGGCGGAATACTTTACGCGTTAGCTAAGG 360
TgCkUg8        AAACAGTTTCACCCTTGCGAGCGTACTACTCAGGCGGAATACTTTACGCGTTAGCTAAGG 360

TgCkUg8-XI-30  TACAAAAAAGAAGATCAACGGCGAAAGCATCTTACTGGGTCGATACTAACACTGAGATAC 420
TgCkUg8        TACAAAAAAGAAGATCAACGGCGAAAGCATCTTACTGGGTCGATACTAACACTGAGATAC 420

TgCkUg8-XI-30  GAAGGTTAAGGTCCATCTTCTGATGCCTTTGTGTATAGACCGAGGGTTCCTGCGTGACGA 480
TgCkUg8        GAAGGTTAAGGTCCATCTTCTGATGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480

TgCkUg8-XI-30  CCTCATTACCATTCACAAACCTTTCAGTCCTCCCCGCCACTCTTCCCGCTGTCGTTCTCT 540
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNNNGTCCTCCCCGCCACTCTTCCCGCTGTCGTTCTCT 540

TgCkUg8-XI-30  CAAATGCCTCCA 552
TgCkUg8        CAAATGCCTCCA 552
```

```
TgCkUg8-X-26   CGCGCTCTCTTTCTTTACTATAGCAGTTGATTCGATACCTCTCATTTCGGAAGCTAAACC 60
TgCkUg8        CGCGCTCTCTTTCTTTACTATAGCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60

TgCkUg8-X-26   GGAACAGCGGATAATCACCCATCCACTGACTACGTTTCGAGTCATGAACTGTTGACGAAT 120
TgCkUg8        NNNNNNNNNGATAATCACCCATCCACTGACTACGTTTCGAGTCATGAACTGTTGACGAAT 120

TgCkUg8-X-26   TAGCATTCGTGTATTTAAAGCTCGAATCTCAGAAAGTAAACTTAGATTAATAAGAGATGA 180
TgCkUg8        TAGCATTCGTGTATTTAAAGCTCGAATCTCAGAAAGTAAACTTAAATTAATAAGAGATGA 180

TgCkUg8-X-26   CATAAATACTAACAAACCACCGATCTGGGATGGAAATACTTTTAACACCGCATAATATGC 240
TgCkUg8        CATAAATACTAACAAACCACCGGTCTGGGATGGAAATACTTTTAACACCGCATAATATGC 240

TgCkUg8-X-26   TATAAAGTACCATTCAGGTAGGCTATGAAGTGGTGTCACGAACCGAGTATATTTGAAAAA 300
TgCkUg8        TATAAAGTACCATTCAGGTAGGCTATGAAGTGGTGTCACGAACCGANNNNNNNNNNNNNN 300

TgCkUg8-X-26   GCAACATTTATATACAAGCTGTACACATATCATGATATTCACTTAGGTAGTCTCCTTCCT 360
TgCkUg8        NNAACATTTATATACAAGCTGTACACATATCATGATATTCACTTAAGTAGTCTCCTTCCT 360

TgCkUg8-X-26   AATGTTAGTCTGTACGGAATAAACGGATCGGATTCTTGTTGG 402
TgCkUg8        AATGTTAGTCTGTACGGAATAAACGGATCGGATTCTTGTTGG 402
```

```
TgCkUg8-IX-24  AATAGCTGCGCCTCAATACATTTCGGGGAGAACCAGGTAGCTCGAAAAATTCGACAAGCA 60
TgCkUg8        AATAGCTGCGCCTCAATACATTTCGGGGAGAACCAGGTAGCTCGAAAAATTCGACAAGCA 60

TgCkUg8-IX-24  TTTCACAACTAACCTAATTTTTCTGCAAAAAAGAATTTTACGGTTCAAGTCCGTTTATCT 120
TgCkUg8        TTTCACAACTAACCTAATTTTTCTGCAAAAAAGAATTTTACGGTTCAAGTCCGTTTATCT 120

TgCkUg8-IX-24  CCATAGATTTATTTTTCCCTTAGATGTTTTAAAGGTTCAGGAAAGGATTACGAGTTTGAT 180
TgCkUg8        CCATAGATTTATTTTTCCCTTAGATGTTTTAAAGGTTCAGGAAAGGATTACGAGTTTGAT 180

TgCkUg8-IX-24  CCTGGCTCCGGAAGAACGCGGATTTCAGATTAAGAAATATTTATAGCGGGGTAGAGCAGG 240
TgCkUg8        CCTGGCTCCGGAAGAACGCGGATTTCAGATTAAGAAATATTTATAGCGGGGTAGAGCAGG 240

TgCkUg8-IX-24  TTGGTAGGTCGTCGGGCTACTTCTAATTCTAATT 274
TgCkUg8        TTGGTAGGTCGTCGGGCTACTTCTAATTCTAATT 274
```

```
TgCkUg8-IX-22  CGTTGCATTAGATTTAGGCGACTCTACAACCAGCCTCTTGACCCACGGGAGCAACCTCGG 60
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCTTGACCCACGGGAGCAACCTCGG 60

TgCkUg8-IX-22  CGTGCGGACGACCTGCCTTCTGTGCAGGGGTGACGGCACACATGCATGGTGTCTCTCGTT 120
TgCkUg8        CGTGCGGACGACCTGCCTTCTGTGCAGGGGTGACGGCACACATGCATGGTGTCTCTCGTT 120

TgCkUg8-IX-22  GGGATTAGTAGCGTCGCGTCTGCATGAAGACCAGCGAGCGACAGCTCATTTCTGCGTCAT 180
TgCkUg8        GGGATTAGTAGCGTCGCGTCTGCATGAAGACCAGCGAGCGACAGCTCATTTCTGCGTCAT 180

TgCkUg8-IX-22  CGGCCACTGGATGCGCACAGGGAGGGAGGGGCCGCATGTGCCACTGCGCCGCAGTATGTG 240
TgCkUg8        CGGCCACTGGATGCGCACAGGGAGGGAGGGGCCGCATGTGCCACTGCGCCGCAGTATGTG 240

TgCkUg8-IX-22  GATTCGAAGCGGGGCCAGGTGAGGCTGCCGCTGCAAGGGTAGGTAGTCGCAAATTTTTGA 300
TgCkUg8        GATTCGAAGCGGGGCCAGGTGAGGCTGCCGCTGCAAGGGTAGGTAGTCGCAAATTTTTGA 300

TgCkUg8-IX-22  AGTGTTGGCGAG 312
TgCkUg8        AGTGTTGGCGAG 312
```

```
TgCkUg8-VIII-21 CGTCTTGGTCACCTCAGTCGCCGATCAACAGTTGGTCTCTGCATTCTCTTCTTACCATCC 60
TgCkUg8         CGTCTTGGTCACCTCAGTCGCCGATCAACAGTTGGTCTCTGCATTCTCTTCTTACCATCC 60

TgCkUg8-VIII-21 CAGAACTGTACGCACGGTTTAGGAGTGTTTTTGCGGGAGTCTCAATTAGCCATATTCATA 120
TgCkUg8         CAGAACTGTACGCACGGTTTAGGAGTGTTTTTGCGGGAGTCTCAATTAGCCATATTCATA 120

TgCkUg8-VIII-21 GTATGGGACAGGTCAATATGAAGTTACGGTGCGCAGGGAGCCTCCCCCTGTCTCGAGCAA 180
TgCkUg8         GTATGGGACAGGTCAATATGAAGTTACGGTGCGCAGGGAGCCTCCCCCTGTCTCGAGCAA 180

TgCkUg8-VIII-21 CCAAGAATTTGTTCGGCACCGCATTGCCAAGAGTAAGTCGAACAAATCTAAAGAGCTCTT 240
TgCkUg8         CCAAGAATTTGTTCGGCACCGCATTGCCAAGAGTAAGTCGAACAAATCTAAAGAGCTCTT 240

TgCkUg8-VIII-21 GCCCTGTTCTGCCGCCTCAAAAAACTAACCTACGACTCGGTCCACCAACGGAGCGCCACG 300
TgCkUg8         GCCCTGTTCTGCCGCCTCAAAAAACTAACCTACGACTCGGTCCACCAACGGAGCGCCACG 300

TgCkUg8-VIII-21 TTCTCGGAATACTTCCTGTTTAGAAGCTTCAGGAGGAAATACTCTGTCGACACTACACAT 360
TgCkUg8         TTCTCGGAATACTTCCTGTTTAGAAGCTTCAGGAGGAAATACTCTGTCGACACTACACAT 360

TgCkUg8-VIII-21 ACAAAACTCCCACATGCAGAACATTCGATTGTTTATGAGTGCAGAACGAGCTTCGGAGGA 420
TgCkUg8         ACAAAACTCCCACATGCAGAACATTCGATTGTTTATGAGTGCAGAACGAGCTTCGGAGGA 420

TgCkUg8-VIII-21 CACACCGTGCTATAAGAACCAAGCACATTGCACACCGGGGGACAAAGCTAGTACTTCATA 480
TgCkUg8         CACACCGTGCTATAAGAACCAAGCACATTGCACACCGGGGGACAAAGCTAGTACTTCATA 480

TgCkUg8-VIII-21 AATCCCAGCAGGGAAATAGGCAGGTGCTACGCATATACAGGTATTGTCGTTCTCAGTGTG 540
TgCkUg8         AATCCCAGCAGGGAAATAGGCAGGTGCTACGCATATACAGGTATTGTCGTTCTCAGTGTG 540

TgCkUg8-VIII-21 GCTGTTTAGATCCTTGCATTAGGTTCACCTGTATGCAGCTCACGACGATTTGCTTAGGAA 600
TgCkUg8         GCTGTTTAGATCCTTGCATTAGGTTCACCTGTATGCAGCTCACGACGATTTGCTTAGGAA 600

TgCkUg8-VIII-21 ACGCGCTTCTGCTTCA 616
TgCkUg8         ACGCGCTTCTGCTTCA 616
```

```
TgCkUg8-VIII-20R  TCTAGCAACCCTGCAGGTATTGATACCGCGCTTAAAGGTGCCTTTTATCCTCATATGTTA 60
TgCkUg8           TCTAGCAACCCTGCAGGTATTGATACCGCGCTTAAAGGTGCCTTTTATCCTCATATGTTA 60

TgCkUg8-VIII-20R  ATGACATGCTAAATGTCTATCACATTTAATTGGATGAATCTTCTTACAAGCGGCTCATTG 120
TgCkUg8           ATGACATGCTAAATGTCTATCACATTTAATTGGATGAATCTTCTTACAAGCGGCTCATTG 120

TgCkUg8-VIII-20R  TATTATGTAGTTTATTCGGAAGATGGAAAACGGAATCGTTATTAAGCGCCAGGTAGTCCT 180
TgCkUg8           TATTATGTAGTTTATTCGGAAGATGGAAAACGGAATCGTTATTAAGCGCCAGGTAGTCCT 180

TgCkUg8-VIII-20R  GGACACTGAATCCATGAACGTGAACATTGTTGTTCTGCTGTATATATTTACACAATTTTA 240
TgCkUg8           GGACACTGAATCCATGAACGTGAACATTGTTGTTCTGCTGTATATATTTACACAATTTTA 240

TgCkUg8-VIII-20R  AATGGTAGATAGGGAACAACCTGTCTCCAAGTCAACAGCTACGCGTCCTCAACAGCCAAG 300
TgCkUg8           AATGGTAGATAGGGAACAACCTGTCTCCAAGTCAACAGCTACGCGTCCTCAACAGCCAAG 300

TgCkUg8-VIII-20R  TGTGCTTTTATAGGTCTCACTCTGTGTATGTCAAAAGCTATCGATCAATCAGCTCTGACG 360
TgCkUg8           TGTGCTTTTATAGGTCTCACTCTGTGTATGTCAAAAGCTATCGATCAATCAGCTCTGACG 360

TgCkUg8-VIII-20R  CTTCTCTCCATATATAGTA 379
TgCkUg8           CTTCTCTCCATATATAGTA 379
```

```
TgCkUg8-VIII-20F  GTCGCTTAGCAAGGTACGCGAGAGAAGAAAAGATCTTCGAGACTGACTAGTACCATACAT 60
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60

TgCkUg8-VIII-20F  AATATTTTTATGATCCTAGGATGGTTTAATAAGTCATAGTTTAGCCGGGAAGTCAGCGTC 120
TgCkUg8           NNNNNNNNNNNNNNNNNAGGATGGTTTAATAAGTCATAGTTTAGCCGGGAAGTCAGCGTC 120

TgCkUg8-VIII-20F  TACCAGATATAACCGAGTATCAACTTAGATGCGCAGATGGACATAATTAATCCTTGTACG 180
TgCkUg8           TACCAGATATAACCGAGTATCAACTTAGATGCGCAGATGGACATAATTAATCCTTGTACG 180

TgCkUg8-VIII-20F  GTTTGTACCTACTTGACGCCTCAGTTTAAGTTAAACAGCCCTTTGTTTACAGCTTGTTCC 240
TgCkUg8           GTTTGTACCTACTTGACGCCTCAGTTTAAGTTAAACAGCCCTTTGTTTACAGCTTGTTCC 240

TgCkUg8-VIII-20F  GTTATATTCAGGAGCATACCGTTATATTTGATGATCTTATGTGTTCACATAGGAATAACT 300
TgCkUg8           GTTATATTCAGGAGCATACCGTTATATTTGATGATCTTATGTGTTCACATAGGAATAACT 300

TgCkUg8-VIII-20F  CAAATCGAAGAAACAGAGACAGTATAGTACCGAGGATACTGAATATGACTGCAGTTATGA 360
TgCkUg8           CAAATCGAAGAAACAGAGACAGTATAGTACCGAGGATACTGAATATGACTGCAGTTATGA 360

TgCkUg8-VIII-20F  GATACAGACAACCAAGTTCTTTATTATTGCTGTGCACCACCGCCCCACTGGACTGCTTAA 420
```

```
TgCkUg8           GATACAGACAACCAAGTTCTTTATTATTGCTGTGCACCACCGCCCCACTGGACTGCTTAA 420

TgCkUg8-VIII-20F  GACAGCTAAAAGTGTTGGATTTCAATATGCTACGACATTAAGGTTATTCCACATCGGTTA 480
TgCkUg8           GACAGCTAAAAGTGTTGGATTTCAATATGCTACGACATTAAGGTTATTCCACATCGGTTA 480

TgCkUg8-VIII-20F  TGTTCTACGCGTAATATATTGATTACT 507
TgCkUg8           TGTTCTACGCGTAATATATTGATTACT 507



TgCkUg8-VIII-19R  CTGTGTCCGCCTCAGCTGTTCGATAAAAATTGCCATTGCATTGGAGCAGAGCGAGGACTT 60
TgCkUg8           CTGTGTCCGCCTCAGCTGTTCGATAAAAATTGCCATTGCATTGGAGCAGAGCGAGGACTT 60

TgCkUg8-VIII-19R  GAGCTTGGGGCCGAGTGCGAGAAAGCATTCTGGGATTCAGGATAATCTAGGGTGAAAGTC 120
TgCkUg8           GAGCTTGGGGCCGAGTGCGAGAAAGCATTCTGGGATTCAGGATAATCTAGGGTGAAAGTC 120

TgCkUg8-VIII-19R  TCTAGATTCACGCAAGCTCTCCGGCGTGTACGCGGGAGGCATGCGCGTTCATCTCCATCT 180
TgCkUg8           TCTAGATTCACGCAAGCTCTCCGGCGTGTACGCGGGAGGCATGCGCGTTCATCTCCATCT 180

TgCkUg8-VIII-19R  TCGCGCCTCTTGCTGGTGCTTCGCGGAGACCTTGTAGCGTGACCGTGTTTGGCAACCACG 240
TgCkUg8           TNNNNNNNNNNNNNGGTGCTTCGCGGAGACCTTGTAGCGTGACCGTGTTTGGCAACCACG 240

TgCkUg8-VIII-19R  TGTGTCGTCCGTCTTCTCATGGGCGGTCCTCACCGCCCAGGTTCACGTCACCGGGGCACA 300
TgCkUg8           TGTGTCGTCCGTCTTCTCATGGGCGGTCCTCACCGCCCAGGTTCACGTCACCGGGGCACA 300

TgCkUg8-VIII-19R  GTGAAAGCAACCTGCAGAACGACGGTACGGGCGAAGGCAGGTTTTTAGCAGCCTACGCGC 360
TgCkUg8           GTGAAAGCAACCTGCAGAACGACGGTACGGGCGAAGGCAGGTTTTTAGCAGCCTACGCGC 360

TgCkUg8-VIII-19R  TGTAAAGGTTGCACTGAGAGA 381
TgCkUg8           TGTAAAGGTTGCACTGAGAGA 381



TgCkUg8-VIII-19F  AGTGAGTGCACAAGGGAGGCGCTTCACACATCCCACGTCGTGCGAAGGCGCGAGCTAGAT 60
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60

TgCkUg8-VIII-19F  GCTTTACCCTGTTTCGCGACTGCGCGCCTTGCGCCAAACGCAGTGAAGGACCAACGACGC 120
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAACGCAGTGAAGGACCAACGACGC 120

TgCkUg8-VIII-19F  GCCTCGGAGTGCCGACATCTGCCCCCGGTAAAATGATACGTGTGTTTCAAAGCCGCCTCT 180
TgCkUg8           GCCTCGGAGTGCCGACATCTGCCCCCGGTAAAATGATACGTGTGTTTCAAAGCCGCCTCT 180

TgCkUg8-VIII-19F  ACATGGCTGATACCGGCAAGAAAATTCGCGTAGGGGGACTTGCCCCTGACGGGCTCCGCG 240
TgCkUg8           ACATGGCTGATACCGGCAAGAAAATTCGCGTAGGGGGACTTGCCCCTGACGGGCTCCGCG 240

TgCkUg8-VIII-19F  TGCCGGAGAAGAACGCGGGCCCGTATACAGAAACGGGACGAATTGCGAGACCGCTGAGAC 300
TgCkUg8           TGCCGGAGAAGAACGCGGGCCCGTATACAGAAACGGGACGAATTGCGAGACCGCTGAGAC 300

TgCkUg8-VIII-19F  GCGCAATCACGAGCATCACCTCCCGTGGGAAACGAGTTCTGCTCTCCGCGTGGATCAATA 360
TgCkUg8           GCGCAATCACGAGCATCACCTCCCGTGGGAAACGAGTTCTGCTCTCCGCGTGGATCAATA 360

TgCkUg8-VIII-19F  CCCCTCAGGCGTTGTCAACTAACCCGTGACCTCACTCGGTCTGTGGGAGCGCTGACCGAT 420
TgCkUg8           CCCCTCAGGCGTTGTCAACTAACCCGTGACCTCACTCGGTCTGTGGGAGCGCTGACCGAT 420

TgCkUg8-VIII-19F  CGATCGAGCGAGAACAACGACGCCGGGAAACGCGCTTCTACTTAGCCCGCCAATGGCGCA 480
TgCkUg8           CGATCGAGCGAGAACAACGACGCCGGGAAACGCGCTTCTACTTAGCCCGCCAATGGCGCA 480

TgCkUg8-VIII-19F  TATTTGCAACGAACATTCCGAGAAGACCGCCAAGATGGGAACGCGTACGCCGTTGAACCT 540
TgCkUg8           TATTTGCAACGAACATTCCGAGAAGACCGCCAAGATGGGAACGCGTACGCCGTTGAACCT 540

TgCkUg8-VIII-19F  GATTGTCTTTCGTAACCTCCACTGTTTCAGCTTTTCGTGCCTCTCCCCCGGGGATAGCGA 600
TgCkUg8           GATTGTCTTTCGTAACCTCCACTGTTTCAGCTTTTCGTGCCTCTCCCCCGGGGATAGCGA 600

TgCkUg8-VIII-19F  AACACTGAACTCCCAGTTTTTTTTGCCTGCCGGATGAGATATCTGAAGACATCAAGGGTC 660
TgCkUg8           AACACTGAACTCCCAGTTTTTTTTGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660

TgCkUg8-VIII-19F  CTGTTGCGCTGCGACTCCGGTTGCCTCAGAGGATATAATGACGACGGCTCTAGCTGCAGA 720
TgCkUg8           NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTAGCTGCAGA 720

TgCkUg8-VIII-19F  CGCGCACC 728
TgCkUg8           CGCGCACC 728



TgCkUg8-VIIb-18   AGACAAAGACCATTCGAGCGAATGTGAAGTTGCAATTGTCTTTATAGGTGGATTTGAAAA 60
TgCkUg8           AGACANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60

TgCkUg8-VIIb-18   ACGGTGAAAGTATGAAGCAATACGAGCCGTTCGTGCTTCTCTGACGATGACGCACACACC 120
```

```
TgCkUg8         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120

TgCkUg8-VIIb-18 GGCACAAGGATGAGTTGAAAAGGACCGCAGAACATCAACAGAACATAAAGATGGCAGGCA 180
TgCkUg8         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180

TgCkUg8-VIIb-18 GGATGCGGCCAGAGGCAAAGGAGCAATGAATGACAAAGAATGGATGCACCTTAATTCCTG 240
TgCkUg8         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTG 240

TgCkUg8-VIIb-18 AAAGCGTTTCCCCTGTAATACTTGTACGTCAGCCCAACCGTTTGTCACGATCTCGTCTCA 300
TgCkUg8         AAAGCGTTTCCCCTGTAATACTTGTACGTCAGCCCAACCGTTTGTCACGATCTCGTCTCA 300

TgCkUg8-VIIb-18 TAGGTAGGAGGCTCGCAAAGCTCCAGAGACAATACGAATCTTGCTGCATGCGCTCAAAAG 360
TgCkUg8         TAGGTAGGAGGCTCGCAAAGCTCCAGAGACAATACGAATCTTGCTGCATGCGCTCAAAAG 360

TgCkUg8-VIIb-18 AGACTTCTTGAATTCTCAAGACCATAAAGATGTGGAACAGAGGGGTCGCGGCACTTCATC 420
TgCkUg8         AGACTTCTTGAATTCTCAAGACCATAAAGATGTGGAACAGAGGGGTCGCGGCACTTCATC 420

TgCkUg8-VIIb-18 ACAAACCCAGCAGCGTGCATGCACCCGAGAATTCGTGC 458
TgCkUg8         ACAAACCCAGCAGCGTGCATGCACCCGAGAATTCGTGC 458


TgCkUg8-VIIb-17 ATCGTCACATCTGGTAAGGTGTTTGTCTACCAGGCATTATAGTACCTAGGATACTGAATA 60
TgCkUg8         ATCGTCACATCTGGTAAGGTGTTTGTCTACCAGGCATTATAGTACCTAGGATACTGAATA 60

TgCkUg8-VIIb-17 TGACTCCAGTTATGAGATACAGACAACCAAGTTCTTTATGATTGCTGTACACCACCCCAC 120
TgCkUg8         TGACTCCAGTTATGAGATACAGACAACCAAGTTCTTTATGATTGCTGTACACCACCCCAC 120

TgCkUg8-VIIb-17 TGGACTGCTTAAGACAGCTAAAAGTGTTGGATGTCAATATCCTACGACATTAAGGTTACT 180
TgCkUg8         TGGACTGCTTAAGACAGCTAAAAGTGTTGGATGTCAATATCCTACGACATTAAGGTTACT 180

TgCkUg8-VIIb-17 CCACATCGGTTATGTTATAGGCGTAATATATGGATTACTGTTATCACTCGTATTAACCGC 240
TgCkUg8         CCACATCGGTTATGTTATAGGCGTAATATATGGATTACTGTTATCACTCGTATTAACCGC 240

TgCkUg8-VIIb-17 GAGAGAAAACTACTACTCAGATGCTAGTCTGATCAGTACCATCGTACTGGGAGTAATCTT 300
TgCkUg8         GAGAGAAAACTACTACTCAGATGCTAGTCTGATCAGTACCATCGTACTGGGAGTAATCTT 300

TgCkUg8-VIIb-17 CTCGGAGACAGGATTATCTGGAAGCTTTCTCTGGGGAGTCAGCTTTTAGCTGCCATGCCA 360
TgCkUg8         CTCGGAGACAGGATTATCTGGAAGCTTTCTCTGGGGAGTCAGCTTTTAGCTGCCATGCCA 360

TgCkUg8-VIIb-17 CACCTTTGGAGAACTGCCTCTCGGCGTCCTGGCGTGACATGTGCCTGTTTGTATATTTAC 420
TgCkUg8         CACCTTTGGAGAACTGCCTCTCGGCGTCCTGGCGTGACATGTGCCTGTTTGTATATTTAC 420

TgCkUg8-VIIb-17 CCAACACAGTGGCTGAGCGAATCCACTCACACCTGCGGCGGGGAAGGAGGAGGAACATCG 480
TgCkUg8         CCAACACAGTGGCTGAGCGAATCCACTCACACCTGCGGCGGGGAAGGAGGAGGAACATCG 480

TgCkUg8-VIIb-17 CTGTCACGGTTGACTCC 497
TgCkUg8         CTGTCACGGTTGACTCC 497


TgCkUg8-VIIa-16 ATTTTGCAAGACGTCAGGGATGTCACCTACACACTTTTTACAATACGGGATTGACGCCCC 60
TgCkUg8         ATTTTGCAAGACGTCAGGGATGTCACCTACACACTTTTTACAATACGGGATTGACGCCCC 60

TgCkUg8-VIIa-16 GCCAAGTGAGCGTTCTGGATTAAAGGTTAAAGGCGGTTAAGAAAAGCTCGCCAGTCGAGC 120
TgCkUg8         GCCAAGTGAGCGTTCTGGATTAAAGGTTAAAGGCGGTTAAGAAAAGCTCGCCAGTCGAGC 120

TgCkUg8-VIIa-16 TCATGCGGTAGTCGCTGCTGCTAAGGTATGAGTAGCAGTGTGCTGCGCGAAAGCCCCAGG 180
TgCkUg8         TCATGCGGTAGTCGCTGCTGCTAAGGTATGAGTAGCAGTGTGCTGCGCGAAAGCCCCAGG 180

TgCkUg8-VIIa-16 AGGCAGAGGCGTGCGAGCGGTGGCACGCGCCTGCTGAAGATAATGCTGGTGGTGCGGCAG 240
TgCkUg8         AGGCAGAGGCGTGCGAGCGGTGGCACGCGCCTGCTGAAGATAATGCTGGTGGTGCGGCAG 240

TgCkUg8-VIIa-16 ACATGTGACTGCTACATGACAAGGTACTCACAACTATTTGACAGGAGGTGGATCATCACA 300
TgCkUg8         ACATGTGACTGCTACATGACAAGGTACTCACAACTATTTGACAGGAGGTGGATCATCACA 300

TgCkUg8-VIIa-16 AAGACTGCACAATAACACGGCGCTGGGGACTGCATTCTTACAGGTCAGCTGTGGCGTGTA 360
TgCkUg8         AAGACTGCACAATAACACGGCGCTGGGGACTGCATTCTTACAGGTCAGCTGTGGCGTGTA 360

TgCkUg8-VIIa-16 GTGTGGGAAGGTGAGGCCGGTTACTGGCTTGTCCGATGCTGGATTATAAACGTGATAGAA 420
TgCkUg8         GTGTGGGAAGGTGAGGCCGGTTACTGGCTTGTCCGATGCTGGATTATAAACGTGATAAAA 420

TgCkUg8-VIIa-16 TATTACCGTTTATGGTTTCGGCCTGAGGGATGTGACCGCTCGAAACTCCAGGCATAGCAT 480
TgCkUg8         TNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480

TgCkUg8-VIIa-16 CCTCTCGGGCGCCCGCCGTCGGTTTCG 507
TgCkUg8         NNNNNNNNNNNNNNNNNNNNNNNNNNN 507
```

```
TgCkUg8-VIIa-15  GTCCTCTTAGTACGTTATAGTATTTACTCACTTGTGTCAGTTTTAGTACGGTCATATCTA 60
TgCkUg8          GTCCTCTTAGTACGTTATAGTATTTACTCACTTGTGTCAGTTTTAGTACGGTCATATCTA 60

TgCkUg8-VIIa-15  TACATACACAGTATCATTTTTCAAGCAAAAATTTTACTTCGAAATAGAAATCATTTATTA 120
TgCkUg8          TACATACACAGTATCATTTTTCAAGCAAAAATTTTACTTCGAAATAGAAATCATTTATTA 120

TgCkUg8-VIIa-15  TATTACCTAGAATGTATATTTAG 143
TgCkUg8          TATTACCTAGAATGTATATTTAG 143



TgCkUg8-VI-14  GACACATGTCTGCGTTTACAAAAGGACTCTCACGGCTTTCCTCTTTGACCTTTTGGATGA 60
TgCkUg8        GACACATGTCTGCGTTTACAAAAGGACTCTCACGGCTTTCCTCTTTGACCTTTTGGATGA 60

TgCkUg8-VI-14  GACGTGACTGCTCAAGACGGAGCAAGCCTCTCTGTTCAACTTCCTTCTGCATCTGAAGAG 120
TgCkUg8        GACGTGACTGCTCAAGACGGAGCAAGCCTCTCTGTTCAACTTCCTTCTGNNNNNNNNNN 120

TgCkUg8-VI-14  ATCGTTTCTGACAAAGCGTTTTTACAGGCAACAGAGCGAGAGGCAAAGTAGCTGAGACGC 180
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAGAGCGAGAGGCAAAGTAGCTGAGACGC 180

TgCkUg8-VI-14  TTGAGACGAAACAGAAGGGGATCTCTCGACTGACAACCCAGGCTAGATGGAACATGCACA 240
TgCkUg8        TTGAGACGAAACAGAAGGGGATCTCTCGACTGACAACCCAGGCTAGATGGAACATGCACA 240

TgCkUg8-VI-14  GATCCTAGTTCTTTGTGAGTGTGTCTGTGCACCACCACTCCACTGGGCTGTTTAGGTAAG 300
TgCkUg8        GATCCTAGTTCTTTGTGAGTGTGTCTGTGCACCACCACTCCACTGGGCTGTTTAGGTAAG 300

TgCkUg8-VI-14  CTAACGGTGTCGGACTCCAATATCTACGACATTAAGGTTACTTCACGTCGGTTACGCTAG 360
TgCkUg8        CTAACGGTGTCGGACTCCAATATCTACGACATTAAGGTTACTTCACGTCGGTTACGCTAG 360

TgCkUg8-VI-14  AAAATACCATGCAGGTACGATACGAAGTGGTGTTACGAACCGGTTGACTGGTATTGAGTT 420
TgCkUg8        AAAATACCATGCAGGTACGATACGAAGTGGTGTTACGAACCGGTTGACTGGTATTGAGTT 420

TgCkUg8-VI-14  ATCTGGATGTGATAGTTCCATCAAGCCAAAAGCCGCTTGTCAGAAGATTAATCCAATTAA 480
TgCkUg8        ATCTGGATGTGATAGTTCCATCAAGCCAAAAGCCGCTTGTCAGAAGATTAATCCAATTAA 480

TgCkUg8-VI-14  ATATGATAGACATTTTGCACCTGTTATTACAATATGATAAACAGCAACTTTGAGCGCTAG 540
TgCkUg8        ATATGATAGACATTTTGCACCTGTTATTACAATATGATAAACAGCAACTTTGAGCGCTAG 540

TgCkUg8-VI-14  ACCCTAGAGAAGGATATCGTTCCACAAGCCGTTTGGGGTTAAGATGATGAAAGAATTGAC 600
TgCkUg8        ACCCTAGAGAAGGATATCGTTCCACAAGCCGTTTGGGGTTAAGATGATGAAAGAATTGAC 600

TgCkUg8-VI-14  ATGTTTTATGTGCTCTAGGAAGCTCTTGATCAGGCGGCTGTAGATAATGCTTTCTTGAAC 660
TgCkUg8        ATGTTTTATGTGCTCTAGGAAGCTCTTGATCAGGCGGCTGTAGATAATGCTTTCTTGAAC 660

TgCkUg8-VI-14  ATTAAACAGTATGGGTTCGTTGTCCCTGGCAACGAAGTCCTCTGTATGCAGATACGAGGC 720
TgCkUg8        ATTAAACAGTATGGGTTCGTTGTCCCTGGCAACGAAGTCCTCTGTATGCAGATACGAGGC 720

TgCkUg8-VI-14  ATTCGACCGAGAGTTGGCGTTATGCATCTGTGAGTACCTGTCTGCAAGTATC 772
TgCkUg8        ATTCGACCGAGAGTTGGCGTTATGCATCTNNNNNNNNNNNNNNNNNNNNN 772



TgCkUg8-VI-13  GAGAGACTCATCACGTGATGTGGCGGCTGGCCGCCCACCACTGAACTTACGGTGCATAGC 60
TgCkUg8        GAGAGACTCATCACGTGATGTGGCGGCTGGCCGCCCACCACTGAACTTACGGTGCATAGC 60

TgCkUg8-VI-13  AGCAGATAAATAACAGGGTACGAACCCTGTCTCATTAGTAGAGACCCGCTACAACAACAG 120
TgCkUg8        AGCAGATAAATAACAGGGTACGAACCCTGTCTCATTAGTAGAGACCCGCTACAACAACAG 120

TgCkUg8-VI-13  GCAGATGCACTGGAGATACGTGAGGCGCTTAAGCGAACTGAAGGAGAGATGCGACGACAC 180
TgCkUg8        GCAGATGCACTGGAGATACGTGAGGCGCTTAAGCGAACTGAAGGAGAGATGCGACGACAC 180

TgCkUg8-VI-13  CGACGACGGCGCTCATGGCTAGACCGGCGCGGACAGCCGAGGAAGCGATGTAGAAGCTAA 240
TgCkUg8        CGACGACGGCGCTCATGGCTAGACCGGCGCGGACAGCCGAGGAAGCGATGTAGAAGCTAA 240

TgCkUg8-VI-13  CCTTGCAGTCCTGAAGTTGTGTGTCTCTGTGGGAAGCGCTGGGTTTTGTCTTGGTGCAAT 300
TgCkUg8        CCTTGCAGTCCTGAAGTTGTGTGTCTCTGTGGGAAGCGCTGGGTTTTGTCTTGGTGCAAT 300

TgCkUg8-VI-13  TGACGTAGAGGTGATGCTCGTCAGAGGGCAAGATGGGAACTGTGAAGTTGTACGTCTTTT 360
TgCkUg8        TGACGTAGAGGTGATGCTCGTCAGAGGGCAAGATGGGAACTGTGAAGTTGTACGTCTTTT 360

TgCkUg8-VI-13  CACTTTCA 368
TgCkUg8        CACTTTCA 368
```

```
TgCkUg8-V-12  AAATGCTCACGTGAAGCTGCCAGGGAGACTGCTGCTGACCGAGGCGGTTTTGCCAGTTTA 60
TgCkUg8       AAATGCTCACGTGAAGCTGCCAGGGAGACTGCTGCTGACCGAGGCGGTTTTGCCAGTTTA 60

TgCkUg8-V-12  GATGGACTGTGTCTATGAGCTGGTGCTGCTGCTGC 95
TgCkUg8       GATGGACTGTGTCTATGAGCTGGTGCTGCTGCTGC 95


TgCkUg8-V-10  CCAGAACACAATTATTTCGTTCAAGCAACTCGCTTACGGAGGAGATCTAGTTACAGCCAG 60
TgCkUg8       CCAGAACACAATTATTTCGTTCAAGCAACTCGCTTACGGAGGAGATCTAGTTACAGCCAG 60

TgCkUg8-V-10  ATAAACCCCGGCAACTATAGCGTCAATGCTTCCTGTGAATACGGCGTCAAACGGAACAAT 120
TgCkUg8       ATAAACCCCGGCAACTATAGCGTCAATGCTTCCTGTGAATACGGCGTCAAACGGAACAAT 120

TgCkUg8-V-10  CTCGCATCCGCTGACAGAATACACTACGCCAACTGGTGTTCTATACGTCTGTAAGACTGT 180
TgCkUg8       CTCGCATCCGCTGACAGAATACACTACGCCAACTGGTGTTCTATACGTCTGTAAGACTGT 180

TgCkUg8-V-10  GACGTCGATACGCACTAGAGAGTCATCGACGCCCGTAAAAGAACTGGACAGGCTGACCCA 240
TgCkUg8       GACGTCGATACGCACTAGAGAGTCATCGACGCCCGTAAAAGAACTGGACAGGCTGACCCA 240

TgCkUg8-V-10  CCACAGAGAGAGTCATGAAACAGAAGCGGGTTTTGTTTCTTGCAAACACCTGCAGGAACC 300
TgCkUg8       CCACAGAGAGAGTCATGAAACAGAAGCGGGTTTTGTTTCTTGCAAACACCTGCAGGAACC 300

TgCkUg8-V-10  AAAGCTGCAATCAGCACTGCGTGCCCGTGGCCGCTTTTATCGGGGATAAAAGCTCGATAG 360
TgCkUg8       AAAGCTGCAATCAGCACTGCGTGCCCGTGGCCGCTTTTATCGGGGATAAAAGCTCGATAG 360

TgCkUg8-V-10  CGTCGACTTGCCTTCGTTTGCTGACTAACGCCCCTCTGGGCCAAGCCTCAGCCCGGTCAG 420
TgCkUg8       CGTCGACTTGCCTTCGTTTGCTGACTAACGCCCCTCTGGGCCAAGCCTCAGCCCGGTCAG 420

TgCkUg8-V-10  GTCCCCCACCGTTGATCTCGTTTCGGATTTCCTGCACCGTCCAATTTATTCACGAAATTA 480
TgCkUg8       GTCCCCCACCGTTGATCTCGTTTCGGATTTCCTGCACCGTCCAATTTATTCACGAAATTA 480

TgCkUg8-V-10  ACTTTGCAGCGCCTGTCGACCGAGTCATCCAATAATATCTTATTGACTACCGCTCGGACA 540
TgCkUg8       ACTTTGCAGCGCCTGTCGACCGAGTCATCCAATAATATCTTATTGACTACCGCTCGGACA 540

TgCkUg8-V-10  ATTGCGCCTGTGCTCTCAGTCCTGTTTGCTAGTCCTCGCTGCCGCTTGGAACGATACCGA 600
TgCkUg8       ATTGCGCCTGTGCTCTCAGTCCTGTTTGCTAGTCCTCGCTGCCGCTTGGAACGATACCGA 600

TgCkUg8-V-10  CTCACTTC 608
TgCkUg8       CTCACTTC 608


TgCkUg8-IV-9  AGGGGAGGAAGGAAACGTAAAATCGGGGCGTGCATGCGTCTCTCTATGCCTCTGTACAGA 60
TgCkUg8       AGGGGAGGAAGGAAACGTAAAATCGGGGCGTGCATGCGTCTCTCTATGCCTCTGTACAGA 60

TgCkUg8-IV-9  GACAGAAGGTGACAGCCTTGCAGATGCCCCATAGGATACTCCTGATACCAAGAACAGCAA 120
TgCkUg8       GACAGAAGGTGACAGCCTTGCAGATGCCCCATAGGATACTCCTGATACCAAGAACAGCAA 120

TgCkUg8-IV-9  GTCGTCCAATTAACCAATCAGTAGCCTCAGGATTTAAAACCATCAAGCTAGTACTTAATG 180
TgCkUg8       GTCGTCCAATTAACCAATCAGTAGCCTCAGGATTTAAAACCATCAAGCTAGTACTTAATG 180

TgCkUg8-IV-9  GAGGGTACTTTGTCCAACCAGGACCACTACCAAATTCAGCACATATACTTTGAGTTACTA 240
TgCkUg8       GAGGGTACTTTGTCCAACCAGGACCACTACCAAATTCAGCACATATACTTTGAGTTACTA 240

TgCkUg8-IV-9  AAACAGAACCTAATGGTCGTAGGAAATGGGAGATCGTGTTAGTTCTTGGGAAAACAATTT 300
TgCkUg8       AAACAGAACCTAATGGTCGTAGGAAATGGGAGATCGTGTTAGTTCTTGGGAAAACAATTT 300

TgCkUg8-IV-9  CCGAACCACCAATATAAATTGGTACAAAGAAGTTACCACTTCATCCGTACAAAGCCGGCA 360
TgCkUg8       CCGAACCACCAATATAAATTGGTACAAAGAAGTTACCACTTCATCCGTACAAAGCCGGCA 360

TgCkUg8-IV-9  TTAAGAACATAATGATCGTAGCTAGACGGGGGGCTTACGTCTGTTCACGCTACAGCGGTG 420
TgCkUg8       TTAAGAACATAATGATCATAGCTAGACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420

TgCkUg8-IV-9  CAGACACGTCTGCATGCACGCATACAGACGTACACATGCTCGTGTAGGTGTAAGCTGGAT 480
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480

TgCkUg8-IV-9  TTGCGGGATAGAAGGCGCATATCTACGTAGTTGTAGGCTGTCAGAGAACAGAGGAGAATT 540
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 540

TgCkUg8-IV-9  CGATTTCGCTTGAACGGTGCTGGCATGGCGA 571
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNN 571


TgCkUg8-IV-8  CAGCCCCATGCCACGACGTCTCGAAATGGCCCGGAGCCGGACTGTCCTGAACCGTGAGTT 60
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60
```

```
TgCkUg8-IV-8  CCTCGCGAGAATCGTCCCCAGCTGCGTCACCGTGGTCGCACTCATGAATCGTAGTGGGTG 120
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120

TgCkUg8-IV-8  TTCAGTGGAGACCGCGCGACTGCTTAACAACGGAGACTGGGCACGAGCGAAGGCACAGAC 180
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180

TgCkUg8-IV-8  GCGGCACCAGACACACGGTGGACGCATCCAGAGGCGCGTGAAATGGCCGTCGACCGCCCA 240
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240

TgCkUg8-IV-8  TGGTACCTCGCTGGCGAAGACTTGTTGCCACGTTTATCTCTGTGCTGGCGCATCTCCAGA 300
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300

TgCkUg8-IV-8  CGCCGTCGAACGCGCTGGATAAAGCCTTGCGACAACGCACACGCCGGCAGCTACTCCTTT 360
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360

TgCkUg8-IV-8  GCTCGGAACAGCGCCACTGACATCTCCGTGAAACACGTATTCCAGAATCAACGGAGGAAC 420
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420

TgCkUg8-IV-8  GCCGTGAACCCCGCACAATGAATGTAGGATGAATAGCCTATTGCTGTGGCATTTTGTCCG 480
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480

TgCkUg8-IV-8  CGCCTATCGCGACGTTGCCCCGCGACAGAGTCACGTTCTATCCTTGTCGCGAAGACCCGC 540
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 540

TgCkUg8-IV-8  GCCACCTACCGCCTGAGACCCAAAATCCACGGACAGCAGCGGCGCGGCCTGAGTCACACG 600
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600

TgCkUg8-IV-8  GCCGAAGCATCAACGAGGGCCAGAAACAGTGAACCGCCACACCTGCATTCATGCGGTCAC 660
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660

TgCkUg8-IV-8  GTCGTTGTCA 670
TgCkUg8       NNNNNNNNNN 670


TgCkUg8-III-7  GGGCAGCATACTCTGTACATTTCTAAATATTCCACGGTGTCGACGCCGAAACGCAGCTTC 60
TgCkUg8        GGGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60

TgCkUg8-III-7  TGGTGGTACGTGGTGTGGACGGTTAATCGCTCCGAGAGAGAATGGGGGCGCTTTGACAGG 120
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120

TgCkUg8-III-7  CGTGTCACTTCCGAACCTGGTTTGTCTCACGCAGCGGGCTGAATAAATGGCATTCGTCGC 180
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180

TgCkUg8-III-7  TCTTCATTTGCGAGTGCATTTGCTTATCCCACTATTCCCGAAGCTAAAATCTACATTGGA 240
TgCkUg8        NNNNNNNNNNNNNNNNCATTTGCTTATCCCACTATTCCCGAAGCTAAAATCTACATTGGA 240

TgCkUg8-III-7  ACCAGCTGTGCACAGTTAAGCGTGCAGAAGCTTGCTGGAGCCCTTTCGACGAGTGTCAAA 300
TgCkUg8        ACCAGCTGTGCACAGTTAAGCGTGCAGAAGCTTGCTGGAGCCCTTTCGACGAGTGTCAAA 300

TgCkUg8-III-7  GCTGAAAAAGTATATCGAGTTTGGTTACTAAATGGGATTGAAAATGCAACTGTGAGTCAC 360
TgCkUg8        GCTGAAAAAGTATATCGAGTTTGGTTACTAAATGGGATTGAAAATGCAACTGTGAGTCAC 360

TgCkUg8-III-7  AAGGTCAGTGGACTGTTCTTTGGATTGTCACGCGTTGGGTAAAACAACCGCACATGCCAG 420
TgCkUg8        AAGGTCAGTGGACTGTTCTTTGGATTGTCACGCGTTGGGTAAAACAACCGCACATGCCAG 420


TgCkUg8-II-5R  CGTGCTTCTTTTGCGGGAGGGGCGCTAGCTGCTGAGCCGAACACCGGCCATTTGACACCA 60
TgCkUg8        CGTGCTTCTTTTGCGGGAGGGGCGCTAGCTGCTGAGCCGAACACCGGCCATTTGACACCA 60

TgCkUg8-II-5R  CGCACCAAAAACCGAACCGGTGCCCGATGATGCGCGCCCAGCTCCAGCAGCGTAGTACTA 120
TgCkUg8        CGCACCAAAAACCGAACCGGTGCCCGATGATGCGCGCCCAGCTCCAGCAGCGTAGTACTA 120

TgCkUg8-II-5R  GAGAACCGCCATGCTTCGTGATCAAACAAATGTCATTCGCCGAGAATTTCTATTTGGTAC 180
TgCkUg8        GAGAACCGCCATGCTTCGTGATCAAACAAATGTCATTCGCCGAGAATTTCTATTTGGTAC 180

TgCkUg8-II-5R  CTGCCCGCGCATTTGAAAGAATTATTACGACATGCGCGTTCTCGGTTGGTGCGGGCAGCT 240
TgCkUg8        CTGCCCGCGCATTTGAAAGAATTATTACGACATGCNNNNNNNNNNNNNNNNNNNNNNN 240

TgCkUg8-II-5R  CCAGAATCGCGCTTCACCGAAGGTTCTTCCGAGAGGCACAGGATACGGTTTTG 293
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNNNNNCGAGAGGCACAGGATACGGTTTTG 293


TgCkUg8-II-5F  CATCTCAGTGGAACTCGGGGATCACTCTCTGGGACATACTGAAATCAGTTGTGACGAGGC 60
TgCkUg8        CATCTCAGTGGAACTCGGGGATCACTCTCTGGGACATACTGAAATCAGTTGTGACGAGGC 60
```

```
TgCkUg8-II-5F  TCCGCTGTTGCCTATAGGGAAACAATGCCGGGATTGTCGCCGTGTTCGCGCTTCAGTGAG 120
TgCkUg8        TCCGCTGTTGCCTATAGGGAAACAATGCCGGGATTGTCGCCGTGTTCGCGCTTCAGTGAG 120

TgCkUg8-II-5F  GCACTTAACCGTGCCGGAGAAATGTGTCGGTGAAGCAATCTCAGACGAGATCCAGGTTTC 180
TgCkUg8        GCACTTAACCGTGCCGGAGAAATGTGTCGGTGAAGCAATCTCAGACGAGATCCAGGTTTC 180

TgCkUg8-II-5F  AAGCGCCCGCAGCTACCAGTGAACATCGCACGTCTGGAAAACAGCGGAGCCGTCTGCTGG 240
TgCkUg8        AAGCGCCCGCAGCTACCAGTGAACATCGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240

TgCkUg8-II-5F  GGTTCTTTTAGAAACGCTGGTACGTGGCGTTTTCAAACAATAACCACACCGCCGGCCCGT 300
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300

TgCkUg8-II-5F  CTGTTAAAGAACAAAGGTGTGGCACCCAGTTGACCGAAGCGGGAAAGAGTTTCGACGGTG 360
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360

TgCkUg8-II-5F  TCCCGGTGTCCGACGCTGCCGATGTGGGCACCGCCTGTAACCACGAAGGCGAGAGCGTCC 420
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420

TgCkUg8-II-5F  ACAGCTGTCTCTGGCGAAGCATCTGATTCTCTAAATAGTTTCTCTGTCACTCAGGCGAGG 480
TgCkUg8        NNNNNNNNNNNNNNNNNNNNNNNNGATTCTCTAAATAGTTTCTCTGTCACTCAGGCGAGG 480

TgCkUg8-II-5F  GCAACGCCGACTTGCCTGTGAGGAACGTGAAGCTATTGTAACTGGAACACCCAGTGAAGT 540
TgCkUg8        GCAACGCCGACTTGCCTGTGAGGAACGTGAAGCTATTGTAACTGGAACACCCAGTGAAGT 540

TgCkUg8-II-5F  TAGCCTCCAACAATGTGGCCAGGCATTCGCTGCGGTTTCTTCCGATGACGGTGTGTGAAC 600
TgCkUg8        TAGCCTCCAACAATGTGGCCAGGCATTCGCTGCGGTTTCTTCCGATGACGGTGTGTGAAC 600

TgCkUg8-II-5F  ATGGTAACCGTTCACCGTCGGCGAACTTGCTGTGACTAAAGGAGGCAGGTGTGTCAGCAA 660
TgCkUg8        ATGGTAACCGTTCACCGTCGGCGAACTTGCTGTGACTAAAGGAGGCAGGTGTGTCAGCAA 660

TgCkUg8-II-5F  ATAGTTCGGTCCCTCGTTGCGTGTCCGCAGGGGAGCCAGGGCTTTTGCCCTCGCTCAACA 720
TgCkUg8        ATAGTTCGGTCCCTCGTTGCGTGTCCGCAGGGGANNNNNNNNNNNNNNNNNNNNNNNN 720

TgCkUg8-II-5F  ACATCACTTCCGTG 734
TgCkUg8        NNNNNNNNNNNNN 734
```

```
TgCkUg8-II-4  ACACACTATCGCACTTTCCCCACATGCGTTGCTGCGGCAGACTCATCAACTCACTACCTA 60
TgCkUg8       ACACACTATCGCACTTTCCCCACATGCGTTGCTGCGGCAGACTCATCAACTCACTACCTA 60

TgCkUg8-II-4  ACAACTCACAAACTGGCCGGGGAACGTGTCAATGTAGATGGAGGCTTTTCAGAATGTGTC 120
TgCkUg8       ACAACTCACAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120

TgCkUg8-II-4  GTTCGCATGTTGCAGTATTATGCGTAACTCCTGCAACGAGCTTTGGAGCTCCCGTTTGTG 180
TgCkUg8       NTTCGCATGTTGCAGTATTATGCGTAACTCCTGCAACGAGCTTTGGAGCTCCCGTTTGTG 180

TgCkUg8-II-4  TCGGCTGAACTGTCACGGACGGTCCAACGGGCAAGACGGTTTCACTGCTCGACAGTTTTC 240
TgCkUg8       TCGGCTGAACTGTCACGGACGGTCCAACGGGCAAGACGGTTTCACTGCTCGACAGTTTTC 240

TgCkUg8-II-4  AAAGGGGCGCCTTTCTTGGGAGCACCTCATCCATCGGTACGACGTAGACGCAGCTTGTAT 300
TgCkUg8       AAAGGGGCGCCTTTCTTGGGAGCACCTCATCCATCGGTACGACGTAGACGCAGCTTGTAT 300

TgCkUg8-II-4  GTCAAGGAAGAAAACACATTATCTTCTTTGAGTAATGTTTTGTAAGCATGGTTCCACATC 360
TgCkUg8       GTCAAGGAAGAAAACACATTATCTTCTTTGAGTAATGTTTTGTAAGCATGGTTCCACATC 360

TgCkUg8-II-4  TATGCGAGGCAGTGCAAAATTGCGGTCGGAGGTTGAGAGCCGCGGATTGCGTCTCTTAGT 420
TgCkUg8       TATGCGAGGCAGTGCAAAATTGCGGTCGGAGGTTGAGAGCCGCGGATTGCGTCTCTTAGT 420

TgCkUg8-II-4  AAGGTGCCTTTCCAGCCACAAGAGACGCGGGTCAAGTCGGGGTGCTGTCGGTAGGGGATT 480
TgCkUg8       AAGGTGCCTTTCCAGCCACAAGAGACGCGGGTCAAGTCGGGGTGCTGTCGGTAGGGGATT 480

TgCkUg8-II-4  GTCGGGCCGTTTCCACCATGACCGCTTCCCAAACCTCGTCGCTGTGGCTGCTACCGATGT 540
TgCkUg8       GTCGGGCCGTTTCCACCATGACCGCTTCCCAAACCTCGTCGCTGTGGCTGCTACCGATGT 540

TgCkUg8-II-4  TTTTGTGTTGTCTTGCTTTTTTTTTAGCAAAGGTCGTGAAGGCAGCGTTGGAAAAATT 597
TgCkUg8       TTTTGTGTTGTCTTGCTTTTTTTTTAGCAAAGGTCGTGAAGGCAGCGTTGGAAAAATT 597
```

```
TgCkUg8-Ib-3  TCCGCCCTTCTGCCACACAGGACATGAGCCCACGCGTCTCTCCCGCGATTCACACGCGCG 60
TgCkUg8       TCCGCCCTTCTGCCACACAGGACATGAGCCCACGCGTCTCTCCCGCGATTCACACGCGCG 60

TgCkUg8-Ib-3  CCGACGGGAGAAACACTGCCCACTCAAACGGTCGTGGAGGTGCAGAGAAGACTCGCTCCG 120
TgCkUg8       CCGACGGGAGAAACACTGCCCACTCAAACGGTCGTGGAGGTGCAGAGAAGACTCGCTCCG 120

TgCkUg8-Ib-3  CAGCGCAGAGAAACGAGGACGGAACGAACCAACTTCACAAGATTCGGAGCAATACAGGCT 180
TgCkUg8       CAGCGCAGAGAAACGAGGACGGAACGAACCAACTTCACAAGATTCGGAGCAATACAGGCT 180
```

```
TgCkUg8-Ib-3  CGTCTATGAAAACCCGGGGAGCGGGAACTTCGCGCCTCCCCTTTCTCGTCTCCGTCTCGC 240
TgCkUg8       CGTCTATGAAAACCCGGGGAGCGGGAACTTCGCGCCTCCCCTTTCTCGTCTCCGTCTCGC 240

TgCkUg8-Ib-3  CAGTCGAGGTGTCCAGCCTCTCTGTCGACACAAACCTTTTTGCCTCCATACAACCGTAAG 300
TgCkUg8       CAGTCGAGGTGTCCAGCCTCTCTGTCGACACAAACCTTTTTGCCTCCATACAACCGTAAG 300

TgCkUg8-Ib-3  ATCATCGACTGACCACCAAAGACCGCTCGACATGAATACATACATAAAGTTTGAGATACT 360
TgCkUg8       ATCATCGACTGACCACCAAAGACCGCTCGACATGAATACATACATAAAGTTTGAGATACT 360

TgCkUg8-Ib-3  ACACCAAAAGCAGGTAGTAGAATTAGAATATATACTTCAGGATGTCCAGAAAACCAGAAT 420
TgCkUg8       ACACCAAAAGCAGGTAGTAGAATTAGAATATATACTTCAGGATGTCCAGAAAACCAGAAT 420

TgCkUg8-Ib-3  AGATGTTGATACAAGAGGTCGCGACATCCCCTCCTTGCACATACAGGGAAACCTCGCCCC 480
TgCkUg8       AGATGTTGATACAAGACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480

TgCkUg8-Ib-3  CTACGCATACGCACTCCCACCGACGTCGCGCTGTACACCCACTCTCTCAAGAGACTTCCC 540
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 540

TgCkUg8-Ib-3  TG 542
TgCkUg8       NG 542
```

```
TgCkUg8-Ib-2  TTCGAAGGAGACCTTACTAGTGCGAACCGCGTCGCGTCGACGCAGGCAGTCGAAGACTTC 60
TgCkUg8       TTCGAAGGAGACCTTACTAGTGCGAACCGCGTCGCGTCGACGCAGGCAGTCGAAGACTTC 60

TgCkUg8-Ib-2  TCTGGAGATGAATGGGAGATTCCAGAGGAGACTTCGCCAGCAGAGAGGATGTCGGCGAGG 120
TgCkUg8       TCTGGAGATGAATGGGAGATTCCAGAGGAGACTTCGCCAGCAGAGAGGATGTCGGCGAGG 120

TgCkUg8-Ib-2  ACACCAACCCCTCCCGGCGGAACACCGGGAGACTGCTGCGGAGATGGAAATGAGGAAAAA 180
TgCkUg8       ACACCAACCCCTCCCGGCGGAACACCGGGAGACTGCTGCGGAGATGGAAATGAGGAAAAA 180

TgCkUg8-Ib-2  ATCTTCGGTCGATTCGCAGCCCTCCACGCGAAGGAGCAAGCCGTGTGTTAACCTCTGACT 240
TgCkUg8       ATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240

TgCkUg8-Ib-2  CAGGTGCCCAGAGACATGCAGGCGCCTCCTTTGGAGCAGTCGGTCCAAGCGCCGAAGCTG 300
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGCTG 300

TgCkUg8-Ib-2  CCGCAGCGGTCCTGAAGCCAACCTCCAGCCAAGCTACCACCGGAAATGTTGCCCTTTATC 360
TgCkUg8       CCGCAGCGGTCCTGAAGCCAACCTCCAGCCAAGCTACCACCGGAAATGTTGCCCTTTATC 360

TgCkUg8-Ib-2  GCCCCTCATCCGCAATGGTGCCGTCGCTTCGCCTATGACTTTAAAACGCCAGCGAAGAGT 420
TgCkUg8       GCCCCTCATCCGCAATGGTGCCGTCGCTTCGCCTATGACTTTAAAACGCCAGCGAAGAGT 420

TgCkUg8-Ib-2  CTGTCCATGGTGCCGCAACCCGAGCTGTCATTCTACGACGCGGTGGTTGTGGAGCGGCAC 480
TgCkUg8       CTGTCCATGGTGCCGCAACCCGAGCTGTCATTCTACGACGCGGTGGTTGTGGAGCGGCAC 480

TgCkUg8-Ib-2  CGGGTTGCTCCTGACGGAAATTGTCAATTCCGTTCCGTCAGCTACGCGTTGCTAGGCACA 540
TgCkUg8       CGGGTTGCTCCTGACGGAAATTGTCAATTCCGTTCCGTCAGCTACGCGTTGCTAGGCACA 540

TgCkUg8-Ib-2  GAGGATGCCCATGCGGAAATCCGGCAAGAGGTGGCTCACTACCTGAGGGGCAACTTTAAT 600
TgCkUg8       GAGGATGCCCATGCGGAAATCCGGCAAGAGGTGGCTCACTANNNNNNNNNNNNNNNNNN 600

TgCkUg8-Ib-2  CGGCTTGGCTGGCTGATAAACCCGGACACGCTGGAGGAAGACGAGGGGAGAATGGCTCGG 660
TgCkUg8       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGAATGGCTCGG 660

TgCkUg8-Ib-2  CTCGACAAGAAATACAGAGTCAGAATTCCGTACAAGACGTACAAAGGCTATACCCTGGCA 720
TgCkUg8       CTCGACAAGAAATACAGAGTCAGAATTCCGTACAAGACGTACAAAGGCTATACCCTGGCA 720

TgCkUg8-Ib-2  GAAGACGAGCTCAAACTTAATTGGGTTATCAGGCTTGGAGACGCACGATACAGGATCTGG 780
TgCkUg8       GAAGACGAGCTCAAACTTAATTGGGTTATCAGGCTTGGAGACGCACGATACAGGATCTGG 780

TgCkUg8-Ib-2  GGTG 784
TgCkUg8       GGTG 784
```

```
TgCkUg8-IV-641.m0156 GACTTTGGTTATCAACCCTGGTATCTTGTGAATCGCTTTGCGGGGCTTTGTTGGGGACAC 60
TgCkUg8              NNNNNNNNNNNNNNNNNNTGGTATCTTGTGAATCGCTTTGCGGGGCTTTGTTGGGGACAC 60

TgCkUg8-IV-641.m0156 AACCCAAACGGAACTGCTGCTCTGATTCTGGGAAGTCCGCCGCTGGGATTGTCAGTATAG 120
TgCkUg8              AACCCAAACGGAACTGCTGCTCTGATTCTGGGAAGTCCGCCGCTGGGATTGTCAGTATAG 120

TgCkUg8-IV-641.m0156 AGGCATTGCCAGGCGTCGGCGTTTTCCACCAGCTAATCACAAAATTCGGCAGAATATCGA 180
TgCkUg8              AGGCATTGCCAGGCGTCGGCGTTTTCCACCAGCTAATCACAAAATTCGGCAGAATATCGA 180

TgCkUg8-IV-641.m0156 CAAACTTCTTCTTGCAATCATCCAAGTCTTTGTTCTGTGGGTTGCAGTATTGGGCGCTAT 240
```

```
TgCkUg8                CAAACTTCTTCTTGCAATCATCCAAGTCTTTGTTCTGTGGGTTGCAGTATTGGGCGCTAT 240

TgCkUg8-IV-641.m0156 CTGCGGCAGGATTCAAAGTGCCCTCCGACCCACAGTCAATGGTCACCGTGTTCCTTGTTG 300
TgCkUg8                CTGCGGCAGGATTCAAAGTGCCCTCCGACCCACAGTCAATGGTCACCGTGTTCCTTGTTG 300

TgCkUg8-IV-641.m0156 TTGTCATATCCACCTTCAAAGCCTCTGAGTTGCTCTCGGCGCCGTAGGAGCATGTGACGA 360
TgCkUg8                TTGTCATATCCACCTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360

TgCkUg8-IV-641.m0156 CGTTCTCGTCTGCAACAGAAGAGGCTCTTGCTTTCACATTGACTGTGAGCTGGCACACTG 420
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420

TgCkUg8-IV-641.m0156 CGGGATTTGCAGCTGCTTTGCCGGTGGTGGTGTATGTACACCCGACGTCAAAGGTCTTGT 480
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480

TgCkUg8-IV-641.m0156 CGGATATAGGAAGGTCTTCTTCATTCAGCTTCAGCGTCCACTTTTGAGATTGAATGGCTT 540
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGAGATTGAATGGCTT 540

TgCkUg8-IV-641.m0156 TCTGTTCCGCAGTGGTATTCATCCACTGAACAGCGCTTGTCGTCCCAAGTACGTCTTTCA 600
TgCkUg8                TCTGTTCCGCAGTGGTATTCATCCACTGAACAGCGCTTGTCGTCCCAAGTACGTCTTTCA 600

TgCkUg8-IV-641.m0156 ATGGAACCTGGACGCTCGCCGCGGCGCCTCCGTTTTCCCCAAAATTGCATGGTGTTTCAC 660
TgCkUg8                ATGGAACCTGGACGCTCGCCGCGGCGCCTCCGTTTTCCCCAAAATTGCATGGTGTTTCAC 660

TgCkUg8-IV-641.m0156 TCTGTTTCTGTTTTGTAACACACACATTTCCAGCCAGAGGATTAGGAATCGATGAGATGG 720
TgCkUg8                TCTGTTTCTGTTTTGTAACACACACATTTCCAGCCAGAGGATTAGGAATCGATGAGATGG 720

TgCkUg8-IV-641.m0156 TTCCACCGTCTTTAGCGGTACACTCCAGAGTGGCCGTGAGACTTCCTTTTGATAACGTCA 780
TgCkUg8                TTCCACCGTCTTTAGCGGTACACTCCAGAGTGGCCGTGAGACTTCCTTTTGATAACGTCA 780

TgCkUg8-IV-641.m0156 GGCTCTGTTTACTAACTGCAGTGCTAGATCCCTCGCCTGCCGTAGAAAATTCACACTTCG 840
TgCkUg8                GGCTCTGTTTACTAACTGCAGTGCTAGATCCCTCGCCTGCCGTAGAAAATTCACACTTCG 840

TgCkUg8-IV-641.m0156 CATGTGTCCCCGTGACATCTACCTTGGAACTCACGCCGACTTGAGACTTCAAACTGAGAC 900
TgCkUg8                CATGTGTCCCCGTGACATCTACCTTGGAACTCACGCCGACTTGAGACTTCAAACTGAGAC 900

TgCkUg8-IV-641.m0156 TCTGAAGTCCTTCATGCAGACTGTCGGCAAGACCCTGTCCGCTAGAGAGCAACAAAATTC 960
TgCkUg8                TCTGAAGTCCTTCATGCAGACTGTCGGCAAGACCCTGTCCGCTAGAGAGCAACAAAATTC 960

TgCkUg8-IV-641.m0156 CACCGATACAAACTGCCATCAACTTTCGGACTTGGGACTTGAACCCTCCAAGCCGTTGCT 1020
TgCkUg8                CACCGATACAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020

TgCkUg8-IV-641.m0156 GCATATTGCCTGTCCTT 1037
TgCkUg8                NNNNNNNNNNNNNNNNN 1037



TgCkUg8-IX-80.m05038 TTTGTCGAGTATGAAGCGTCGGGTTCGCTTCTTCAACTGCTCCCGGGCGGCACTCGGGTT 60
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNCGCTTCTTCAACTGCTCCCGGGCGGCACTCGGGTT 60

TgCkUg8-IX-80.m05038 GGGGGTCGCGGCTTTGTTCGTCGCTTGTCTCAGTGACATTTCGAGGGCTGTCGCAACAGA 120
TgCkUg8                GGGGGTCGCGGCTTTGTTCGTCGCTTGTCTCAGTGACATTTCGAGGGCTGTCGCAACAGA 120

TgCkUg8-IX-80.m05038 AACGCCGACAATGAGACCCACATGCAAGGAAGAAAAGCAGATGACCACGTGCATGTGCGC 180
TgCkUg8                AACGCCGACAATGAGACCCACATGCAAGGAAGAAAAGCAGATGACCACGTGCATGTGCGC 180

TgCkUg8-IX-80.m05038 AGGTCCTTCTGGTGCAGCCGCGGCAGCGGCCGCGAAAACAATTGCACTTTGTCCCAGTCG 240
TgCkUg8                AGGTCCTTCTGGTGCAGCCGCGGCAGCGGCCGCGAAAACAATTGCACTTTGTCCCAGTCG 240

TgCkUg8-IX-80.m05038 AGTAATGGAATCACTGTTGACTGTTCCACAGATAGTTTCACTGTTGTGCCTTCCGACACA 300
TgCkUg8                AGTAATGGAATCACTGTTGACTGTTCCACAGATAGNNNNNNNNNNNNNNNNNNNNNNNN 300

TgCkUg8-IX-80.m05038 AGTAAGATTTGCATTGGAGAAGGTCATGACGAAACACTGACGGCGTGTAACAGAAACAAA 360
TgCkUg8                NNNNNNNNNTGCATTGGAGAAGGTCATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360

TgCkUg8-IX-80.m05038 GTATCCACAATCAAGGAATTTCTGGGCGTGGCGCACGGCACCAACGCACCACAGTGGACG 420
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420

TgCkUg8-IX-80.m05038 TCCGACCACGCTCTAACAATTCCAACAGAGCAATTCTCCATAATTGACAAAACCTTATTC 480
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480

TgCkUg8-IX-80.m05038 GCAGGATGCTTAAGAAAAAGTGGGACCAACCAGGGCACGAAGGAGTGCTTGGTGCCTGCC 540
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 540

TgCkUg8-IX-80.m05038 AGTGTCAAAGCCAGGGCCTCGGCAGTGAAGAATGGTGTCCTCATTTGCGCCTACGGCACA 600
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600

TgCkUg8-IX-80.m05038 GAGAGCAACAATACTGTTTCAGCGGTAACACTGGACGCCAAGAACAACTCTCTGACAATC 660
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
```

```
TgCkUg8-IX-80.m05038 ATCTGTGACACTGAAGGACAGATGCACCCAGAAAAGGATTCCTTGACTGCTTGCCACTGC 720
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTGCCACTGC 720

TgCkUg8-IX-80.m05038 ACCAATTCCAACACAGAAAACTGTAAGACCTTGAACCTGAGTTCATGCCGAGCTTCACCG 780
TgCkUg8              ACCAATTCCAACACAGAAAACTGTAAGACCTTGAACCTGAGTTCATGCCGAGCTTCACCG 780

TgCkUg8-IX-80.m05038 AGAGGTGGTGGACGCCGGTTGACACAGACAGTAATGCGTTCAAACTCGTGATTCCAGAGG 840
TgCkUg8              AGAGGTGGTGGACGCCGGTTGACACAGACAGTAATGCGTTCAAACTCGTGATTCCAGAGG 840

TgCkUg8-IX-80.m05038 GAGGCTTTCCTGTTGAGGAAGAAAGGATTGTGTTGGGGTGCACTGAAAACAACAGGGTTC 900
TgCkUg8              GAGGCTTTCCTGTTGAGGAAGAAAGGATTGTGTTGGGGTGCACTGAAAACAACAGGGTTC 900

TgCkUg8-IX-80.m05038 AAGCCTTGAACGGTGACGGTGACAACGACACAGCTGTAAAGCTACCGAC 949
TgCkUg8              AAGCCTTGAACGGTGACGGTGACAACGACACAGCTGTAAAGCTACCGAC 949
```

```
TgCkUg8-XII-65.m0116 ATACTTTCTGCATCATCCCTGCCATCATCCCCACCATCGGCCTCCCCAGAATTCGTGTGG 60
TgCkUg8              ATACTTTCTGCATCATCCCTGCCATCATCCCCACCATCGGCCTCCCCAGAATTCGTGTGG 60

TgCkUg8-XII-65.m0116 TCTGTTGATTCGGATGAGCTTGGAATCGAAGGAGTCGATACGGTGGTTTGTGAATCCGAT 120
TgCkUg8              TCTGTTGATTCGGATGAGCTTGGAATCGAAGGAGTCGATACGGTGGTTTGTGAATCCGAT 120

TgCkUg8-XII-65.m0116 CCGGCGGCAGACACCGTTTTGAACCGCATACAAAAGCTCAACCCTACACTCAAGTACTCT 180
TgCkUg8              CCGGCGGCAGACACCGTTTTGAACCGCATACAAAAGCTCAACCCTACACTCAAGTACTCT 180

TgCkUg8-XII-65.m0116 GCTGGAGGATTTTTAAACGTGTACATCACAACCTTTCCCGCGGAGTCACGTCCTCTTTCA 240
TgCkUg8              GCTGGAGGATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240

TgCkUg8-XII-65.m0116 CCAGATTCTGACTCTTCCCGTGTCCACAATTTTGCTGTGTAGCCGGGAAACACATCCTTG 300
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300

TgCkUg8-XII-65.m0116 AATTGGATTGTCTTTGTGAAATCGCACCCTCCATTCATATATTGATATGCAGCGCTGTAA 360
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360

TgCkUg8-XII-65.m0116 GCGTCAGTCAATGTTCCAGATGAATTCTCGCTAGTTTCGGGTTGATAAACGGCACTCGCT 420
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420

TgCkUg8-XII-65.m0116 GTTTCGTCAGTGACCCGGATCGTGCCAGAAGATCCAAGAGAAAAGACTTCTTCAACATCT 480
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480

TgCkUg8-XII-65.m0116 CCCTCCAATCCCCCTTTCGGAATGATGGTGCTAAAATCAACCTCAGCTGTCGGTTGACCG 540
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 540

TgCkUg8-XII-65.m0116 GATTCGGCGGCGACGCACTGCAAGACAGTGCCAATGAAAAGGACTGAACACCCGCCGCGA 600
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCACCCGCCGCGA 600

TgCkUg8-XII-65.m0116 AAGATCTGCCTTGCCATTGCGAGTGCGGATC 631
TgCkUg8              AAGATCTGCCTTGCCATTGCGAGTGCGGATC 631
```

```
TgCkUg8-X-33.m02185  TGATGCTCATTACCGTCGGATCGCTGCTGACCGCTTCCGAATCGGTCCAGCTTTCTGAAG 60
TgCkUg8              TGATGCTCATTACCGTCGGATCGCTGCTGACCGCTTCCGAATCGGTCCAGCTTTCTGAAG 60

TgCkUg8-X-33.m02185  GCATGAAGAGGCTCAGCATGAGAGGAAGAAGTCCCAGCCCGAAAACGGGAAGATTTGAGA 120
TgCkUg8              GCATGAAGAGGCTCAGCATGAGAGGAAGAAGTCCCAGCCCGAAAACGGGAAGATTTGAGA 120

TgCkUg8-X-33.m02185  GTGGAGACGAAGGGACATCTACAATGTCGCCCTCCGTCGCAGCACGCCAACAGGAACTAG 180
TgCkUg8              GTGGAGACGAAGGGACATCTACAATGTCGCCCTCCGTCGCAGCACGCCAACAGGAACTAG 180

TgCkUg8-X-33.m02185  GGCTGCTCCGTCCGGAAGAGAGGCTCATTGCGGGTCAAGCCAAAGCAGCGGCGCTACAGA 240
TgCkUg8              GGCTGCTCCGTCCGGAAGAGAGGCTCATTGCGGGTCAAGCCAAAGCAGCGGCGCTACAGA 240

TgCkUg8-X-33.m02185  CTGTCCATCAATTAGGAGCAGTTGTTCTTACCCCAGAGCAAGCTAAAGCCGCATTGCTAG 300
TgCkUg8              CTGTCCATCAATTAGGAGCAGTTGTTCTTACCCCAGAGCAAGCTAAAGCCGCATTGCTAG 300

TgCkUg8-X-33.m02185  ACGAAATTCTCAGAGCCACGCAAAATTTGGACCTCAAGAAGTACGAGAATCTCAACACGG 360
TgCkUg8              ACGAAATTCTCAGAGCCACGCAAAATTTGGACCTCAAGAAGTACGAGAATCTCAACACGG 360

TgCkUg8-X-33.m02185  AGCAACAAAAGGCCTATGAACAAGTCCAGAAGGATCTTTCGCAACTGAGTCCGGAAACGA 420
TgCkUg8              AGCAACAAAAGGCCTATGAACAAGTCCAGAAGGATCTTTCGCAACTGAGTCCGGAAACGA 420

TgCkUg8-X-33.m02185  AGGCTTTGCTCATTGAAAATCATAGGAAGGAGAAATCCCTACTTGAACAGGCGAAAAGGC 480
TgCkUg8              AGGCTTTGCTCATTGAAAATCATAGGAAGGAGAAATCCCTACTTGAACAGGCGAAAAGGC 480

TgCkUg8-X-33.m02185  TGTTCCGAAAACGCCACTACCACGTTACCAGGCAGGCGGCTTTGGCTGGACAAATCCTCA 540
```

```
TgCkUg8                TGTTCCGAAAACGCCACTACCACGTTACCAGGCAGGCGGCTTTGGCTGGACAAATCCTCA 540

TgCkUg8-X-33.m02185    ACGAGCAGAGGGATGCGAGTGGAGCATTACAATCAGGTGCTGTGAAGGCTGCGATCCGAA 600
TgCkUg8                ACGAGCAGAGGGATGCGAGTGGAGCATTACAATCAGGTGCTGTGAAGGCTGCGATCCGAA 600

TgCkUg8-X-33.m02185    AAGCCAACGAGCAGTACAACGTCGCTGAAGAAGACAAGAACTTTAACGAGGAACAGCATG 660
TgCkUg8                AAGCCAACGAGCAGTACAACGTCGCTGAAGAAGACAAGAACTTTAACGAGGAACAGCATG 660

TgCkUg8-X-33.m02185    CAGCACAACTGAAAAAGGTAGGCGCTATGCCCTAGTCGTAAAAACGACGACGAACCCGCG 720
TgCkUg8                CAGCACAACTGAAAAAGGTAGGCGCTATGCCCTNNNNNNNNNNNNNNNNNCGAACCCGCG 720

TgCkUg8-X-33.m02185    TCGGAAATCCCGAATCTGTTTGGGATGGCTTTGACCGGTATGAAGAGGTCAAACACAGCT 780
TgCkUg8                TCGGAAATCCCGAATCTGTTTGGGATGGCTTTGACCGGTATGAAGAGGTCAAACACAGCT 780

TgCkUg8-X-33.m02185    TCTTGTCAGACCAACATTGGTCAATTCTGATTTTTTCCGTTTATTCGTTTTCCATGATGA 840
TgCkUg8                TCTTGTCAGACCAACATTGGTCAATTCTGATTTTTTCCGTTTATTCGTTTTCCATGATGA 840

TgCkUg8-X-33.m02185    ACAATCACTTTGAATATGGTCTGCCGCTCAGCACGAA 877
TgCkUg8                ACAATCACTTTGAATATGGTCTGCCGCTCAGCACGAA 877



TgCkUg8-VIIa-20.m038   CGGTCTCCAAGGCTTGTAGCGGGAGCATCCGTTTCTGAGGCTCTCGATTCAAGAGGTTTC 60
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60

TgCkUg8-VIIa-20.m038   GGATGAGGTCTTGAACCAGCTCAGGCGTGGAAGGACAATCTGCGAAGTGTAAGTAGTCCC 120
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120

TgCkUg8-VIIa-20.m038   AGATGCCGTCGGCCGGAGTTGGACGTTCCTTGCACCAGATGCAGTACAAAGTTATACCGA 180
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGA 180

TgCkUg8-VIIa-20.m038   GTTGCCACGCGTCAGTGGTGAATGTATAGGTGATGTCTGTAGTCTGAAACGGTCGCTCCG 240
TgCkUg8                GTTGCCACGCGTCAGTGGTGAATGTATAGGTGATGTCTGTAGTCTGAAACGGTCGCTCCG 240

TgCkUg8-VIIa-20.m038   GAGGCTCGTAACCGGGAGTACCTATCGCGGGTCCAACCGAATTATTGATTCTATACGTTC 300
TgCkUg8                GAGGCTCGTAACCGGGAGTACCTATCGCGGGTCCAACCGAATTATTGATTCTATACGTTC 300

TgCkUg8-VIIa-20.m038   CGAAGTCGCCGAGAAACAGGCGACCGTCTTTCAAGAGGAGGAAATTCGCCGGTTTGATAT 360
TgCkUg8                CGAAGTCGCCGAGAAACAGGCGACCGTCTTTCAAGAGGAGGAAATTCGCCGGTTTGATAT 360

TgCkUg8-VIIa-20.m038   CCGTATGCACAATTCCCTGAGCTTGAACATTGGCCACTAGTCTGATTGCCTGACTGGATA 420
TgCkUg8                CCGTATGCACAATTCCCTGAGCTTGAACATTGGCCACTAGTCTGATTGCCTGACTGGATA 420

TgCkUg8-VIIa-20.m038   GGTACATTCGAACGACCAGGCCTAATTCACTGTTATTGACAGACGCATCTCCAAATACCC 480
TgCkUg8                GGTACATTCGAACGACCAGGCCTAATTCACTGTTATTGACAGACGCATCTCCAAATACCC 480

TgCkUg8-VIIa-20.m038   ATGAAATGACTTTGCTCATGTCCGTTTCTGCCCGCATCATGAGAAGAAAATAGTTTGGTA 540
TgCkUg8                ATGAAATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 540

TgCkUg8-VIIa-20.m038   CCCACCGAGTCGTCAAACCAATCACGACCTCTGTGGATGCTGGCTGTCCCTCTAACATCA 600
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600

TgCkUg8-VIIa-20.m038   CAACATCACTAGGAACCATGAATCTACAGCGTTCCTGGGCATCCTTCGCCGTCTTGGCTA 660
TgCkUg8                NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGTCTTGGCTA 660

TgCkUg8-VIIa-20.m038   GACTAAAGTTCCTGTAGCAGAACGACTCCCTCTGCAAGTCACGCATAGTCTCATCGGTGG 720
TgCkUg8                GACTAAAGTTCCTGTAGCAGAACGACTCCCTCTGCAAGTCACGCATAGTCTCATCGGTGG 720

TgCkUg8-VIIa-20.m038   GCTCCTTTTCTGACATGAAAACCTTAACAGCCAACTCTTCATTCGTCTCCACGTCTGTAG 780
TgCkUg8                GCTCCTTTTCTGACATGAAAACCTTAACAGCCAACTCTTCATTCGTCTCCACGTCTGTAG 780

TgCkUg8-VIIa-20.m038   CCTCATATACAGTGGCGAATCCACCAGAGCCAAGGGGTGCACCTCTCACCAGCGTCCGGG 840
TgCkUg8                CCTCATATACAGTGGCGAATCCACCAGAGCCAAGGGGTGCACCTCTCACCAGCGTCCGGG 840

TgCkUg8-VIIa-20.m038   TTTCACCGGTAGTCACAGAAACAAATCTTTCTGGAACGTCCGGAGGCCAGGCAGCGACAG 900
TgCkUg8                TTTCACCGGTAGTCACAGAAACAAATCTTTCTGGAACGTCCGGAGGCCAGGCAGCGACAG 900

TgCkUg8-VIIa-20.m038   TCAGAATTCTATCCAGTCGTTCACTATTTCCTAGCGCTTGAGGCTGCCCGATTTGTTCTT 960
TgCkUg8                TCAGAATTCTATCCAGTCGTTCACTATTTCCTAGCGCTTGAGGCTGCCCGATTTGTTCTT 960

TgCkUg8-VIIa-20.m038   GCATGTGCGCCAGGATGCGTTTGCCGATGACGCATCCAGAATCCGCCTTCTCAAAAAACA 1020
TgCkUg8                GCATGTGCGCCAGGATGCGTTTGCCGATGACGCATCCAGAATCCGCCTTCTCAAAAAACA 1020

TgCkUg8-VIIa-20.m038   ATTCAGATCGCCGTCTTTGTGCCCGCAAACTTCGCAGTCGGTTACGCCCTTGAGAAAAGT 1080
TgCkUg8                ATTCAGATCGCCGTCTTTGTGCCCGCAAACTTCGCAGTCGGTTACGCCCTTGAGAAAAGT 1080

TgCkUg8-VIIa-20.m038   ACCTCGGGATTCCGCGTGTGAAGAACGATTTCGCTTTACGCCAAAGTTGAGCAAGCCTCT 1140
TgCkUg8                ACCTCGGGATTCCGCGTGTGAAGAACGATTTCGCTTTACGCCAANNNNNNNNNNNNNNNN 1140
```

```
TgCkUg8-VIIa-20.m038 GTCGTACGGACAGACGGGGTCTCGAGGAGCTATCAGCTGCGGAACCCGAGACTTCACCAT 1200
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1200

TgCkUg8-VIIa-20.m038 CTCTTTCTCTCCGTCTCAAGAGATTCAGAAGCCTCTGAAAAATCGTGCGCCGTTGGTTTG 1260
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1260

TgCkUg8-VIIa-20.m038 TTTCTTGCGCATTTACCCGGTGTTGAGCCCTTTCCTCCAGCAATGAAACGTCTCGAGTGC 1320
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1320

TgCkUg8-VIIa-20.m038 TTTCTGTCGCTCCTGCTAAATATTTGTAGTGGCCAACTGTAGCAGGCCACCGCTTGTCAG 1380
TgCkUg8              NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1380

TgCkUg8-VIIa-20.m038 CAACGTGCTGTTGCGAATCCAAACTTGTCGGTTTGGAGTCGAGTTTTGAAGGATCAAGTG 1440
TgCkUg8              NNNNNNGCTGTTGCGAATCCAAACTTGTCGGTTTGGAGTCGAGTTTTGAAGGATCAAGTG 1440

TgCkUg8-VIIa-20.m038 TGATTCCGGTCCCATCCTGGACTTGGAAGAGCAGGAAGACAAGCGCTACATTTAACACCG 1500
TgCkUg8              TGATTCCGGTCCCATCCTGGACTTGGAAGAGCAGGAAGACAAGCGCTACATTTAACACCG 1500

TgCkUg8-VIIa-20.m038 CAAGACAGGCTGTCTTCGGGAGAAGAGTCGCTAAACCCATTCGGACGACGGTACGCGTAA 1560
TgCkUg8              CAAGACAGGCTGTCTTCGGGAGAAGAGTCGCTAAACCCATTCGGACGACGGTACGCGTAA 1560

TgCkUg8-VIIa-20.m038 GAGGTGGCCGCTGTACCGAAAACATCACAACTTTCACACAA 1601
TgCkUg8              GAGGTGGCCGCTGTACCGAAAACATCACAACTTTCACACAA 1601
```

TgCkUg9:

```
TgCkUg9 SAG1 TGTAGTCACTGCTGATTCTCGCTGTTCTCGGCAAGGGCTGACGACCGGAGTACAGTTTTTGTGGGCAGAGCCGCTGTGCA 80
TgCkUg9      TGTAGTCACTGCTGATTCTCGCTGTTCTCGGCAAGGGCTGACGACCGGAGTACAGTTTTTGTGGGCAGAGCCGCTGTGCA 80

TgCkUg9 SAG1 GCTTTCCGTTGTTCTCGGTTGTGTCACATGTGTCATTGTCGTGTAAACACACGGTTGTATGTCGGTTTCGCTGCACCACT 160
TgCkUg9      GCTTTCCGTTGTTCTCGGTTGTGTCACATGTGTCATTGTCGTGTAAACACACGGTTGTATGTCGGTTTCGCTGCACCACT 160

TgCkUg9 SAG1 TCATTATTTCTTCTGGTTTTTTGGCGAGTATGTTTCCGAAGGCAGTGAGACGCGCCGTCACGGCAGGGGTGTTTGCCGCG 240
TgCkUg9      TCATTATTTCTTCTGGTTTTTTGGCGAGTATGTTTCCGAAGGCAGTGAGACGCGCCGTCACGGCAGGGGTGTTTGCCGCG 240

TgCkUg9 SAG1 CCCACACTGATGTCGTTCTTGCGATGTGGCGCTATGGCATCGGATCCCCCTCTTGTTGCCAATCAAGTTGTCACCTGCCC 320
TgCkUg9      CCCACACTGATGTCGTTCTTGCGATGTGGCGCTATGGCATCGGATCCCCCTCTTGTTGCCAATCAAGTTGTCACCTGCCC 320

TgCkUg9 SAG1 AGATAAAAAATCGACAGCCGCGGTCATT 348
TgCkUg9      AGATAAAAAATCGACAGCCGCGGTCATT 348
```

```
TgCkUg9 SAG2 AGCGCTGCTTGCGATTCTGTGTGTTGTTTCACGGGTTGCAGTTCTAGGAACTGAGTTGTGATTGTGCACAATTGCGGTGT 80
TgCkUg9      AGCGCTGCTTGCGATTCTGTGTGTTGTTTCACGGGTTGCAGTTCTAGGAACTGAGTTGTGATTGTGCACAATTGCGGTGT 80

TgCkUg9 SAG2 GACACCTTCTGTCTCGTTCCAATCTTTGTCTTGTCGGAACTATGAGTTTCTCAAAGACCACGAGCCTAGCGTCGCTAGCG 160
TgCkUg9      GACACCTTCTGTCTCGTTCCAATCTTTGTCTTGTCGGAACTATGAGTTTCTCAAAGACCACGAGCCTAGCGTCGCTAGCG 160

TgCkUg9 SAG2 CTCACGGGCTTGTTTGTTGTGTTCAAGTTCGCTCTTGCGTCCACCACCGAGACGCCAGCGCCCATTGAGTGCACTGCCGG 240
TgCkUg9      CTCACGGGCTTGTTTGTTGTGTTCAAGTTCGCTCTTGCGTCCACCACCGAGACGCCAGCGCCCATTGAGTGCACTGCCGG 240

TgCkUg9 SAG2 CGCAACGAAGACTGTTGAGGCACCCTCCAGTGGTTCCGTTGTCTTCCAATGTGGGGATAAACTAACCATCAGTCCCAGTG 320
TgCkUg9      CGCAACGAAGACTGTTGAGGCACCCTCCAGTGGTTCCGTTGTCTTCCAATGTGGGGATAAACTAACCATCAGTCCCAGTG 320

TgCkUg9 SAG2 GCGAAGGTGATGTCTTTTATGGCAAGGAATGCACAGACTCGAGGAAGTTGACGACTGTCCTTCCAGGTGCGGTCTTGAAA 400
TgCkUg9      GCGAAGGTGATGTCTTTTATGGCAAGGAATGCACAGACTCGAGGAAGTTGACGACTGTCCTTCCAGGTGCGGTCTTGAAA 400

TgCkUg9 SAG2 GCTAAGGTCGAGCAGCCCCCGAAAGGTCCTGCTACCTACACACTGTCTTACGACGGTACTCCCGAGAAACCTCAGGTTCT 480
TgCkUg9      GCTAAGGTCGAGCAGCCCCCGAAAGGTCCTGCTACCTACACACTGTCTTACGACGGTACTCCCGAGAAACCTCAGGTTCT 480

TgCkUg9 SAG2 CTGTTACAAGTGCGTTGCCGAAGCAG 506
TgCkUg9      CTGTTACAAGTGCGTTGCCGAAGCAG 506
```

```
TgCkUg9 BTUB CGGCATTGATCCGGTGAGGGAGAACGGGTTCCGGAGTGCGGGTTCCCGGACGCTGTCTTTTGTCCTGTATTCTGCGTGAC 80
TgCkUg9      CGGCATTGATCCGGTGAGGGAGAACGGGTTCCGGAGTGCGGGTTCCCGGACGCTGTCTTTTGTCCTGTATTCTGCGTGAC 80

TgCkUg9 BTUB AGCTTCGCCAGTGTAAATGGAGCCGTTTTCCCTGTGGAGAATCGCGGAGAATGGAGGACTTTTCCGAGCTCGCAGGTGCC 160
TgCkUg9      AGCTTCGCCAGTGTAAATGGAGCCGTTTTCCCTGTGGAGAATCGCGGAGAATGGAGGACTTTTCCGAGCTCGCAGGTGCC 160
```

```
TgCkUg9 BTUB ACCCTCCACCTCGACGCAACCAAGTGCACGCACATTTGCCGGGTGGTGACCCTAAGACACCGCAGGTCTACCTGCGCGTT 240
TgCkUg9      ACCCTCCACCTCGACGCAACCAAGTGCACGCACATTTGCCGGGTGGTGACCCTAAGACACCGCAGGTCTACCTGCGCGTT 240

TgCkUg9 BTUB TTTTTCAGTCTTTGCACACAGTTGCACCGAAAGTCATGTTTTTTGCGAAGACATCGGTCGTTCTGGTGGGGGAATACTTC 320
TgCkUg9      TTTTTCAGTCTTTGCACACAGTTGCACCGAAAGTCATGTTTTTTGCGAAGACATCGGTCGTTCTGGTGGGGGAATACTTC 320

TgCkUg9 BTUB ACTCCTGTTCGCGCCTATGTGCGCAGACAGGTGTCCACCCTCCGCATACGG 371
TgCkUg9      ACTCCTGTTCGCGCCTATGTGCGCAGACAGGTGTCCACCCTCCGCATACGG 371


TgCkUg9 GRA6 GGGTCGCTTTTTTGAAACAGCAGGAAAACAGCTTCGTGGTGCCACGTAGCGTGCTTGTTGGCGACTACCTTTTTTTCTTG 80
TgCkUg9      GGGTCGCTTTTTTGAAACAGCAGGAAAACAGCTTCGTGGTGCCACGTAGCGTGCTTGTTGGCGACTACCTTTTTTTCTTG 80

TgCkUg9 GRA6 GGAGTGTCGGCGAAATGGCACACGGTGGCATCTATCTGAGGCAGAAGCGTAACTTCTGTCCTTTAACTGTCTCCACAGTT 160
TgCkUg9      GGAGTGTCGGCGAAATGGCACACGGTGGCATCTATCTGAGGCAGAAGCGTAACTTCTGTCCTTTAACTGTCTCCACAGTT 160

TgCkUg9 GRA6 GCTGTGGTCTTTGTAGTCTTCATGGGTGTACTCGTCAATTCGTTGGGTGGAGTCGCTGTCGCAGCAGACAGCGGTGGTGT 240
TgCkUg9      GCTGTGGTCTTTGTAGTCTTCATGGGTGTACTCGTCAATTCGTTGGGTGGAGTCGCTGTCGCAGCAGACAGCGGTGGTGT 240

TgCkUg9 GRA6 TAGGCAGACCCCTTCGGAAACCGGTTCGAGCGGTGGACAGCAAGAAGCAGTGGGGACCACTGAAGA 306
TgCkUg9      TAGGCAGACCCCTTCGGAAACCGGTTCGAGCGGTGGACAGCAAGAAGCAGTGGGGACCACTGAAGA 306


TgCkUg9 C22-8 TTTCTCGTTCCTTTCGCGCTGCGTTTCGTTTCTTTTTCAAAAACGGACGGCATGCGGTTCGCATTCCCGCCGTGAAGCAA 80
TgCkUg9       TTTCTCGTTCCTTTCGCGCTGCGTTTCGTTTCTTTTTCAAAAACGGACGGCATGCGGTTCGCATTCCCGCCGTGAAGCAA 80

TgCkUg9 C22-8 TCGAAGAATTTCCGAAGTCAAGAAGAGCGAACCTTCTGTCATCTCCCCTGCCTTTCTCTCAGATCCCTTCGCTTCTTTGT 160
TgCkUg9       TCGAAGAATTTCCGAAGTCAAGAAGAGCGAACCTTCTGTCATCTCCCCTGCCTTTCTCTCAGATCCCTTCGCTTCTTTGT 160

TgCkUg9 C22-8 GCAGGGTTTTGGAGAGATTGAAATTCCTCGAGAAGCAAATTCGGAGGTGCCACTGGAAGGCAACCGCTGCCTCCCACAAC 240
TgCkUg9       GCAGGGTTTTGGAGAGATTGAAATTCCTCGAGAAGCAAATTCGGAGGTGCCACTGGAAGGCAACCGCTGCCTCCCACAAC 240

TgCkUg9 C22-8 ACTGGAAACGCACTGGAACTTTTCAGCATTTTCGCTGAACGGAAAATGCAGCAAACCTGTGAACTCGCACAGCTGTAGCG 320
TgCkUg9       ACTGGAAACGCACTGGAACTTTTCAGCATTTTCGCTGAACGGAAAATGCAGCAAACCTGTGAACTCGCACAGCTGTAGCG 320

TgCkUg9 C22-8 ACGAGCACTACCCCGAAAAACGACCCTCGAGACTGCATGCATGCAGTTCATGAGGAACCGGAAAGGAAAAAGACACTTTC 400
TgCkUg9       ACGAGCACTACCCCGAAAAACGACCCTCGAGACTGCATGCATGCAGTTCATGAGGAACCGGAAAGGAAAAAGACACTTTC 400

TgCkUg9 C22-8 CCCGATCCTTCTCTACGCATCAAAGTCGAATTGAATCTAAAAAGGAGAGAAATCGATTTGCTTCAAACTCTGAGTTCCAA 480
TgCkUg9       CCCGATCCTTCTCTACGCATCAAAGTCGAATTGAATCTAAAAAGGAGAGAAATCGATTTGCTTCAAACTCTGAGTTCCAA 480

TgCkUg9 C22-8 GTTCT 485
TgCkUg9       GTTCT 485


TgCkUg9 C29-2 TTTTTCCGGAAGCGGAGCATGTGGACACGTATGCCATTTCAAATGAATCGAGGACCGGAGGGCTAACGAGACGCCAGTGT 80
TgCkUg9       TTTTTCCGGAAGCGGAGCATGTGGACACGTATGCCATTTCAAATGAATCGAGGACCGGAGGGCTAACGAGACGCCAGTGT 80

TgCkUg9 C29-2 ACACACTGGAGGAGGTCCGTCCATTCCTGTTCCTCGTTTTCCGATTCTGTGTGTCTTTCCCATTTTCGCGCTCTTCCACT 160
TgCkUg9       ACACACTGGAGGAGGTCCGTCCATTCCTGTTCCTCGTTTTCCGATTCTGTGTGTCTTTCCCATTTTCGCGCTCTTCCACT 160

TgCkUg9 C29-2 TCTATGCGTTTATTTGATTGCCTTCCAGCCATGCTGTTCAACAAGAGTGGCTTTTCGCGGGAGTAAGAACTCAGCCACAT 240
TgCkUg9       TCTATGCGTTTATTTGATTGCCTTCCAGCCATGCTGTTCAACAAGAGTGGCTTTTCGCGGGAGTAAGAACTCAGCCACAT 240

TgCkUg9 C29-2 GCAGTACACAAAAGTCCCTCGAATGCTCTCTGAACACACCGAACTCCCGGGAAGGTGCTCACAAACATGCCCACCACCGG 320
TgCkUg9       GCAGTACACAAAAGTCCCTCGAATGCTCTCTGAACACACCGAACTCCCGGGAAGGTGCTCACAAACATGCCCACCACCGG 320

TgCkUg9 C29-2 ATATCCCGTTGTGCTAAAGTGTTTCATCAACTTCTCTCTCTTGATATCCATTCTCACATCACTCCTACGTATTCTGCTCA 400
TgCkUg9       ATATCCCGTTGTGCTAAAGTGTTTCATCAACTTCTCTCTCTTGATATCCATTCTCACATCACTCCTACGTATTCTGCTCA 400

TgCkUg9 C29-2 GTCCGAACAG 410
TgCkUg9       GTCCGAACAG 410


TgCkUg9 L358 CGCAGCATGCCGCGAAGAATGGAATGTTCTCTTCCTCGACACGGGCGGAGAAAGCGAAACCTTCCTGCTGCGCGGTTTCC 80
TgCkUg9      CGCAGCATGCCGCGAAGAATGGAATGTTCTCTTCCTCGACACGGGCGGAGAAAGCGAAACCTTCCTGCTGCGCGGTTTCC 80

TgCkUg9 L358 GTCGCTGTCTCGTCGCTCGCTTCTCTCCCCGCTTCTCCGTCCGAAACGAGTCGCGAGAGAGGCGGCAGGCGAATGCAGGA 160
TgCkUg9      GTCGCTGTCTCGTCGCTCGCTTCTCTCCCCGCTTCTCCGTCCGAAACGAGTCGCGAGAGAGGCGGCAGGCGAATGCAGGA 160

TgCkUg9 L358 GAAACTGATCTCTGCCTCTCTCCTCGCCTTCTTGTCTATCAGCGGGGTGGCCGCGCGACATTGCAGCGTCGACGCGCAGC 240
TgCkUg9      GAAACTGATCTCTGCCTCTCTCCTCGCCTTCTTGTCTATCAGCGGGGTGGCCGCGCGACATTGCAGCGTCGACGCGCAGC 240

TgCkUg9 L358 CGAGCGAAGGAAGCGGGGAAGACGAAGAGGCTCCGTTCCCCCGACAGGACGCATCCGACAAGCCGAAGGCAGAAAGAGGA 320
TgCkUg9      CGAGCGAAGGAAGCGGGGAAGACGAAGAGGCTCCGTTCCCCCGACAGGACGCATCCGACAAGCCGAAGGCAGAAAGAGGA 320

TgCkUg9 L358 CATGCATGCGAGTGTGTCGTCCGGAAGAGCAGAGACTCGCACCCTTGCAGGGCCAAGACGC 381
TgCkUg9      CATGCATGCGAGTGTGTCGTCCGGAAGAGCAGAGACTCGCACCCTTGCAGGGCCAAGACGC 381
```

```
TgCkUg9 PK1  TCTTCTTCATGGAGTAAAAAACTGTTTGGAATGTGTCGGCAGCGACTCTAATCATCTGCTGCCAAATGGCGAATAACGAG 80
TgCkUg9      TCTTCTTCATGGAGTAAAAAACTGTTTGGAATGTGTCGGCAGCGACTCTAATCATCTGCTGCCAAATGGCGAATAACGAG 80

TgCkUg9 PK1  TGCATGTGACGTATGTTCATCAGTACGTGAAGAACCGTTCTCCAGAACAAGCTGTCTCGTAGAGTGGTCAGTACGTTACC 160
TgCkUg9      TGCATGTGACGTATGTTCATCAGTACGTGAAGAACCGTTCTCCAGAACAAGCTGTCTCGTAGAGTGGTCAGTACGTTACC 160

TgCkUg9 PK1  TGGATCTCGACGAAAACGCGGTCAGCATCGGTGATCTCGTTGAATGATCTTTTGTTTATAAGCTTGATTGCAACGATCTC 240
TgCkUg9      TGGATCTCGACGAAAACGCGGTCAGCATCGGTGATCTCGTTGAATGATCTTTTGTTTATAAGCTTGATTGCAACGATCTC 240

TgCkUg9 PK1  TCCTGGTGGTGACAAAGTGCACGCATTTCGCCACTTTGTTGTCGATGACACAAAAAGAGAGAGTTTTAAGGAAACCCATG 320
TgCkUg9      TCCTGGTGGTGACAAAGTGCACGCATTTCGCCACTTTGTTGTCGATGACACAAAAAGAGAGAGTTTTAAGGAAACCCATG 320

TgCkUg9 PK1  GTTTTGCATCACGGCGAAAACGTGTGCCCAACGAACGCGAAACTGGTACCAGGCCACCAAACATTCGATCCCTGTATTCC 400
TgCkUg9      GTTTTGCATCACGGCGAAAACGTGTGCCCAACGAACGCGAAACTGGTACCAGGCCACCAAACATTCGATCCCTGTATTCC 400

TgCkUg9 PK1  GTACAGATTGCACCTTAACGCAGTTCAATGTGGATCGCCTTATACAAGTTGGCTATAGATCTCAACTAGCAAAACTCTCT 480
TgCkUg9      GTACAGATTGCACCTTAACGCAGTTCAATGTGGATCGCCTTATACAAGTTGGCTATAGATCTCAACTAGCAAAACTCTCT 480

TgCkUg9 PK1  CTTTCAGACCTAATACAGGTGTACTTTGGTTTACACTAGAGAAAATTGTGCATGACCTGCACGTGGCAAATTCAGGTGCG 560
TgCkUg9      CTTTCAGACCTAATACAGGTGTACTTTGGTTTACACTAGAGAAAATTGTGCATGACCTGCACGTGGCAAATTCAGGTGCG 560

TgCkUg9 PK1  GAAACAGTCTGGACAAAATCGTCTTCCATTGTGCCCTACTGCTTATGAAAAGCATATGGGCAACAAAAAAGCCCCTGATA 640
TgCkUg9      GAAACAGTCTGGACAAAATCGTCTTCCATTGTGCCCTACTGCTTATGAAAAGCATATGGGCAACAAAAAAGCCCCTGATA 640

TgCkUg9 PK1  CTCGGGTCAGCATGTGACACTGTTGGGATTATGTTGCGCAGTTAATTATTCAGCGAAATGCGGGAAGAACGACACCCATA 720
TgCkUg9      CTCGGGTCAGCATGTGACACTGTTGGGATTATGTTGCGCAGTTAATTATTCAGCGAAATGCGGGAAGAACGACACCCATA 720

TgCkUg9 PK1  CAAATTCTGCATCTTAACTGTCCTACACATAAGGTATTGCGCTCCGAACACGTTGCCACTCGCATCGCAGTAACACTCTC 800
TgCkUg9      CAAATTCTGCATCTTAACTGTCCTACACATAAGGTATTGCGCTCCGAACACGTTGCCACTCGCATCGCAGTAACACTCTC 800

TgCkUg9 PK1  GCAGGGAAACCCACTAATGCATCCTATTCGAGACGC 836
TgCkUg9      GCAGGGAAACCCACTAATGCATCCTATTCGAGACGC 836
```

```
TgCkUg9-XI-UPRT1  TTGTCGATCCCCGATATTCGACAAACGACCAGGAAGAAAGCATTCTCCAGGACATCATCA 60
TgCkUg9           TTGTCGATCCCCGATATTCGACAAACGACCAGGAAGAAAGCATTCTCCAGGACATCATCA 60

TgCkUg9-XI-UPRT1  CGAGGTAATCCTTCAACCGAAGTTTGCTTTCCGTGACTCTGCCTGTTGGTTATACTGCGT 120
TgCkUg9           CGAGGTAATCCTTCAACCGAAGTTTGCTTTCCGTGACTCTGCCTGTTGGTTATACTGCGT 120

TgCkUg9-XI-UPRT1  GGCCTTCCCGTCCTGCGGCCCCCTTTCCTCCGCTTGCTGTTTAAATGCTCGTCCTCGTTT 180
TgCkUg9           GGCCTTCCCGTCCTGCGGCCCCCTTTCCTCCGCTTGCTGTTTAAATGCTCGTCCTCGTTT 180

TgCkUg9-XI-UPRT1  TCCTTCCTGCCGCATCCCCGTATATTTTAAGGAGAGGGAAACAGGCGTGAGTTGGACGGC 240
TgCkUg9           TCCTTCCTGCCGCATCCCCGTATATTTTAAGGAGAGGGAAACAGGCGTGAGTTGGACGGC 240

TgCkUg9-XI-UPRT1  ATGAAAGTTCTCGGCCTGTATGCCGGTTGTGGCGGTCGTTTGCAGATTGCTTTTTTCTTC 300
TgCkUg9           ATGAAAGTTCTCGGCCTGTATGCCGGTTGTGGCGGTCGTTTGCAGATTGCTTTTTTCTTC 300

TgCkUg9-XI-UPRT1  GAATCGGTGCTGTAACCCTCGCGAAGAACGACGCTGCAAACGACTTCTCGAACTCTCAGT 360
TgCkUg9           GAATCGGTGCTGTAACCCTCGCGAAGAACGACGCTGCAAACGACTTCTCGAACTCTCAGT 360

TgCkUg9-XI-UPRT1  CGTGTACTTTACGTGCTTCCTTTCAGGGACCTCCCCCCCGCGTTACTCATTTGTATTCACA 420
TgCkUg9           CGTGTACTTTACGTGCTTCCTTTCAGGGACCTCCCCCCCGCGTTACTCATTTGTATTCACA 420

TgCkUg9-XI-UPRT1  GCTACGAAGTGTCTTGCAAGGTGGATTTCTGCCAGGCTCCATGTCTCACTCGTTGCGTTT 480
TgCkUg9           GCTACGAAGTGTCTTGCAAGGTGGATTTCTGCCAGGCTCCATGTCTCACTCGTTGCGTTT 480

TgCkUg9-XI-UPRT1  TCGGAAAAGTTCATTGTGAACGTTCCCCTTGCGTGTCATGACTTTATCAGGTTTCCCAAT 540
TgCkUg9           TCGGAAAAGTTCATTGTGAACGTTCCCCTTGCGTGTCATGACTTTATCAGGTTTCCCAAT 540

TgCkUg9-XI-UPRT1  GTGGTGCTCATGAAGCAGACGGCTCAGCTTCGAGCGATGATGACCATCATTCGTGGTGAG 600
TgCkUg9           GTGGTGCTCATGAAGCAGACGGCTCAGCTTCGAGCGATGATGACCATCATTCGTGGTGAG 600

TgCkUg9-XI-UPRT1  TGGTGACAGGGTAGAACAGACAGACGAGGAAGGAGAAAAAAAAGGAATCAAGCCACTAAGC 660
TgCkUg9           TGGTGACAGGGTAGAACAGACAGACGAGGAAGGAGAAAAAAAAGGAATCAAGCCACTAAGC 660

TgCkUg9-XI-UPRT1  GGCGACACACTGTTGGTGCCTGAAGCCT 688
TgCkUg9           GGCGACACACTGTTGGTGCCTGAAGCCT 688
```

```
TgCkUg9-X-26  CGCGCTCTCTTTCTTTACTATAGCAGTTGATTCGATACCTCTCATTTCGGAAGCTAAACC 60
TgCkUg9       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60

TgCkUg9-X-26  GGAACAGCGGATAATCACCCATCCACTGACTACGTTTCGAGTCATGAACTGTTGACGAAT 120
TgCkUg9       NNNNNNNNNGATAATCACCCATCCACTGACTACGTTTCGAGTCATGAACTGTTGACGAAT 120

TgCkUg9-X-26  TAGCATTCGTGTATTTAAAGCTCGAATCTCAGAAAGTAAACTTAGATTAATAAGAGATGA 180
TgCkUg9       TAGCATTCGTGTATTTAAAGCTCGAATCTCAGAAAGTAAACTTAGATTAATAAGAGATGA 180

TgCkUg9-X-26  CATAAATACTAACAAACCACCGATCTGGGATGGAAATACTTTTAACACCGCATAATATGC 240
TgCkUg9       CATAAATACTAACAAACCACCGATCTGGGATGGAAATACTTTTAACACCGCATAATATGC 240

TgCkUg9-X-26  TATAAAGTACCATTCAGGTAGGCTATGAAGTGGTGTCACGAACCGAGTATATTTGAAAAA 300
TgCkUg9       TATAAAGTACCATTCAGGTAGGCTATGAAGTGGTGTCACGAACCGANNNNNNNNNNNNN 300

TgCkUg9-X-26  GCAACATTTATATACAAGCTGTACACATATCATGATATTCACTTAGGTAGTCTCCTTCCT 360
TgCkUg9       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCT 360

TgCkUg9-X-26  AATGTTAGTCTGTACGGAATAAACGGATCGGATTCTTGTTGGCCTGGCACCTGTTTAGTA 420
TgCkUg9       AATGTTAGTCTGTACGGAATAAACGGATCGGATTCTTGTTGGCCTGGCACCTGTTTAGTA 420

TgCkUg9-X-26  ACTGGATGAACGCTTTTTGCGCCTGGTATGTATGGATAATACTCTACTCTTCTATAATTT 480
TgCkUg9       ACTGGATGAACGCTTTTTGCGCCTGGTATGTATGGATAATACTCTACTCTTCTATAATTT 480

TgCkUg9-X-26  AACCGCTACTGCTGGGACTGTATATTATGTACTTACGGTAGTACTATCAAGCCTCTTCTC 540
TgCkUg9       AACCGCTACTGCTGGGACTGTATATTATGTACTTACGGTAGTACTATCAAGCCTCTTCTC 540

TgCkUg9-X-26  CCAAACAGATTTCATGGAAAACCCGAAATTCGCAGGTTTGGTTGACATTTAGCCGCTAAT 600
TgCkUg9       CCAAACAGATTTCATGGAAAACCCGAAATTCGCAGGTTTGGTTGACATTTAGCCGCTAAT 600

TgCkUg9-X-26  ATACAATCATGTAAGATATATTTATCTATCGCACGTTCGGTCTAAAGTCCCGTTATACTG 660
TgCkUg9       ATACAATCATGTAAGATATATTTATCTATCGNNNNNNNNNNNNNNNNNNNNNNNNNNN 660

TgCkUg9-X-26  CATAGATCACATAGCTTCTATTTACTCTAGGAGGTACAGCAGGTGTCGTTACGGGTAATG 720
TgCkUg9       NNNNNNNNNNATAGCTTCTATTTACTCTAGGAGGTACAGCAGGTGTCGTTACGGGTAATG 720

TgCkUg9-X-26  CAGGTATGGATATCGCACTACACGGTACATATTATATTGCTGCACATTTCCACTTTGTAC 780
TgCkUg9       CAGGTATGGATATCGCACTACACGGTACATATTATATTGCTGCACATTTCCACTTTGTAC 780

TgCkUg9-X-26  TATCTTGTGGAGCTATCTTAGCAACTATAT 810
TgCkUg9       TATCTTGTGGAGCTATCTTAGCAACTATAT 810
```

```
TgCkUg9-VIIa-16 ATTTTGCAGACGTCAGGGATGTCACCTACACACTTTTTACAATACGGGATTGACAGCTCC 60
TgCkUg9         ATTTTGCAGACGTCAGGGATGTCACCTACACACTTTTTACAATACGGGATTGACAGCTCC 60

TgCkUg9-VIIa-16 CGCCAAGTGAGCGTTCTGGATTAAAGGTTAAAGGCGGTTAAGAAAAGCTCGCCAGTCGAG 120
TgCkUg9         CGCCAAGTGAGCGTTCTGGATTAAAGGTTAAAGGCGGTTAAGAAAAGCTCGCCAGTCGAG 120

TgCkUg9-VIIa-16 CTCATGCGGTAGTCGCTGCTGCTAAGGTATGAGTAGCAGTGTGTTGCGCGAAAGCCCCAG 180
TgCkUg9         CTCATGCGGTAGTCGCTGCTGCTAAGGTATGAGTAGCAGTGTGTTGCGCGAAAGCCCCAG 180

TgCkUg9-VIIa-16 GAGGCAGAGGCGTGCGAGCGGTGGCACGCGCCTGCTGAAGATAATGCTGGTGGTGCGGCA 240
TgCkUg9         GAGGCAGAGGCGTGCGAGCGGTGGCACGCGCCTGCTGAAGATAATGCTGGTGGTGCGGCA 240

TgCkUg9-VIIa-16 GACATGTGCCTGCTACATGACAAGGTACTCACAACTATTTGACAGGAGGTGGATCATCAC 300
TgCkUg9         GACATGTGCCTGCTACATGACAAGGTACTCACAACTATTTGACANNNNNNNNNNNNNNN 300

TgCkUg9-VIIa-16 AAAGACTGCACAATAACACGGCGCTGGGGACTGCATTCTTACAGGTCAGCTGTGGCGTGT 360
TgCkUg9         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACAGGTCAGCTGTGGCGTGT 360

TgCkUg9-VIIa-16 AGTGTGGGAAGGTGAGGCCGGTTAGTGGCTTGTCCGATGCTGGATTATAAACGTGATAAA 420
TgCkUg9         AGTGTGGGAAGGTGAGGCCGGTTAGTGGCTTGTCCGATGCTGGATTATAAACGTGATAAA 420

TgCkUg9-VIIa-16 ATATTACCGTTTATGGTTTCGGCCTGAGGGATGTGACCGCTCGAAACTCCAGGCATAGCA 480
TgCkUg9         ATATTACCGTTTATGGTTTCGGCCTGAGGGATGTGACCGCTCGAAACTCCAGGCATAGCA 480

TgCkUg9-VIIa-16 TCCTCGCGGGCGTCCGCCGTCGGTTTCG 508
TgCkUg9         TCCTCGCGGGCGTCCGCCGTCGGTTTCG 508
```

```
TgCkUg9-VI-13  GAGAGACTCATCACGTGATGTGGCGGCTGGCCGCCCACCACTGAACTTACGGTGCATAGC 60
TgCkUg9        GAGAGACTCATCACGTGATGTGGCGGCTGGCCGCCCACCACTGAACTTACGGTGCATAGC 60

TgCkUg9-VI-13  AGCAGATAAATAACAGGGTACGAACCCTGTCTCATTAGTAGAGACCCGCTACAACAACAG 120
TgCkUg9        AGCAGATAAATAACAGGGTACGAACCCTGTCTCATTAGTAGAGACCCGCTACAACAACAG 120
```

```
TgCkUg9-VI-13  GCAGATGCACTGGAGATACGTGAGGCGCTTAAGCGAACTGAAGGAGAGATGCGACGACAC 180
TgCkUg9        GCAGATGCACTGGAGATACGTGAGGCGCTTAAGCGAACTGAAGGAGAGATGCGACGACAC 180

TgCkUg9-VI-13  CGACGACGGCGCTCATGGCTAGACCGGCGCGGACAGCCGAGGAAGCGATGTAGAAGCTAA 240
TgCkUg9        CGACGACGGCGCTCATGGCTAGACCGGCGCGGACAGCCGAGGAAGCGATGTAGAAGCTAA 240

TgCkUg9-VI-13  CCTTGCAGACCTGAAGTTGTGTGTCTCTGTGGGAAGCGCTGGGTTTTGTCTTGGTGCAAT 300
TgCkUg9        CCTTGCAGACCTGAAGTTGTGTGTCTCTGTGGGAAGCGCTGGGTTTTGTCTTGGTGCAAT 300

TgCkUg9-VI-13  TGACGTAGAGGTGATGCTCGTCAGAGGGCAAGATGGGAACTGTGAAGTTGTACGTCTTTT 360
TgCkUg9        TGACGTAGAGGTGATGCTCGTCAGAGGGCAAGATGGGAACTGTGAAGTTGTACGTCTTTT 360

TgCkUg9-VI-13  CACTTTCA 368
TgCkUg9        CACTTTCA 368


TgCkUg9-IV-8   CAGCCCCATGCCACGACGTCTCGAAATGGCCCGGAGCCGGACTGTCCTGAACCGTGAGTT 60
TgCkUg9        CAGCCCCATGCCACGACGTCTCGAAATGGCCCGGAGCCGGACTGTCCTGAACCGTGAGTT 60

TgCkUg9-IV-8   CCTCGCGAGAATCGTCCCCAGCTGCGTCACCGTGGTCGCACTCATGAATCGTAGTGGGTG 120
TgCkUg9        CCTCGCGAGAATCGTCCCCAGCTGCGTCACCGTGGTCGCACTCATGAATCGTAGTGGGTG 120

TgCkUg9-IV-8   TTCAGTGGAGACCGCGCGACTGCTTAACAACGGAGACTGGGCACGAGCGAAGGCACAGAC 180
TgCkUg9        TTCAGTGGAGACCGCGCGACTGCTTAACAACGGAGACTGGGCACGAGCGAAGGCACAGAC 180

TgCkUg9-IV-8   GCGGCACCAGACACACGGTGGACGCATCCAGAGGCGCGTGAAATGGCCGTCGACCGCCCA 240
TgCkUg9        GCGGCACCAGACACACGGTGGACGCATCCAGAGGCGCGTGAAATGGCCGTCGACCGCCCA 240

TgCkUg9-IV-8   TGGTATCTCGCTGGCGACGTCATGTTGCCACGTTTATCTCTGTGCTGGTGTATCTCGAGA 300
TgCkUg9        TGGTATCTCGCTGGCGACGTCATGTTGCCACGTTTATCTCTGTGCTGGTGTATCTCGAGA 300

TgCkUg9-IV-8   CGCCGTGGAACGCGCTGGATAAAGACGTGCGACGGCAAACAGGCCGGCTGCTACTCCTTT 360
TgCkUg9        CGCCGTGGAACGCGCTGGATAAAGACGTGCGACGGCAAACAGGCCGGCTGCTACTCCTTT 360

TgCkUg9-IV-8   GCTCGGAACAGCGCCACTGACATCTCCGTGAAACACGTATTCCAGAATCAACGGAGGAAC 420
TgCkUg9        GCTCGGAACAGCGCCACTGACATCTCCGTGAAACACGTATTCCAGAATCAACGGAGGAAC 420

TgCkUg9-IV-8   GCCGTGAACCCCGCACAATGAATGTAGGATGAATAGCCTATTGCTGTGGCATTTTGTCCG 480
TgCkUg9        GCCGTGAACCCCGCACAATGAATGTAGGATGAATAGCCTATTGCTGTGGCATTTTGTCCG 480

TgCkUg9-IV-8   CGCCTATCGCGACGTTGCCCCGCGACAGAGTCACGTTCTATCCTTGTCGCGAAGACCCGC 540
TgCkUg9        CGCCTATCGCGACGTTGCCCCGCGACAGAGTCACGTTCTATCCTTGTCGCGAAGACCCGC 540

TgCkUg9-IV-8   GCCACCTACCGCCTGAGACCCAAAATCCACGGACAGCAGCGGCGCGGCCTGAGTCACACG 600
TgCkUg9        GCCACCTACCGCCTGAGACCCAAAATCCACGGACAGCAGCGGCGCGGCCTGAGTCACACG 600

TgCkUg9-IV-8   GCCGAAGCATCAACGAGGGCCAGAAACAGTGAACCGCCACACCTGCATTCATGCGGTCAC 660
TgCkUg9        GCCGAAGCATCAACGAGGGCCAGAAACAGTGAACCGCCACACCTGCATTCATGCGGTCAC 660

TgCkUg9-IV-8   GTCGTTGTCA 670
TgCkUg9        GTCGTTGTCA 670


TgCkUg9-II-5R  CCACAATGCCCCGTCCGCCGCGTGATTCTCCCTGTGTCAACGCCTATGCCCACAGAGACA 60
TgCkUg9        CCACAATGCCCCGTCCGCCGCGTGATTCTCCCTGTGTCAACGCCTATGCCCACAGAGACA 60

TgCkUg9-II-5R  ATGTCGTCGTAGGGAAAGGTTTTCGTCGCGTAGTCCGTAAGAAAGACAAGCAGCGTTGTG 120
TgCkUg9        ATGTCGTCGTAGGGAAAGGTTTTCGTCGCGTAGTCCGTAAGAAAACAAGCAGCGTTGTG 120

TgCkUg9-II-5R  TAAAGCTCGTCAAGTGCTTGTGAGTCCTGTGGCAGGAACCGCAGCCCCATGAAGAGAAAG 180
TgCkUg9        TAAAGCTCGTCAAGTGCTTGTGAGTCCTGTGGCAGGAACCGCAGCCCCATGAAGAGAAAG 180

TgCkUg9-II-5R  CACACTCGCACCACACGTCAGTTTATGTCAACTCCGGCTTCACACCTCCATCCGCTTTCC 240
TgCkUg9        CACACTCGCACCACACGTCAGTTTATGTCAACTCCGGCTTCACACCTCCATCCGCTTTCC 240

TgCkUg9-II-5R  GGGACACCAGACTCACGCCGCTGCAAATGGAGTGGTCGTGCCACGAGGTCGCCGTGTGGG 300
TgCkUg9        GGGACACCAGACTCACGCCGCTGCAAATGGAGTGGTCGTGCCACGAGGTCGCCGTGTGGG 300

TgCkUg9-II-5R  ACCGTGGCAGGATGACAAACAACATTCTGTTCGGCTCCTGATACCGGAGGGCGCCGCAGA 360
TgCkUg9        ACCGTGGCAGGATGACAAACAACATTCTGTTCGGCTCCTGATACCGGAGGGCGCCGCAGA 360

TgCkUg9-II-5R  CTGCGGGGCGAAATCGTGAATAGCCCTTGCTTCAGCGGAGGGTCGTTATGTGGGCACCAG 420
TgCkUg9        CTGCGGGGCGAAATCGTGAATAGCCCTTGCTTCAGCGGAGGGTCGTTATGTGGGCACCAG 420

TgCkUg9-II-5R  GGCCAAACTTGCGCGTGGATCCGTGCTTCGTCTGCGGGAGGGGCGCTAGCTGCTGAGCCG 480
```

```
TgCkUg9         GGCCAAACTTGCGCGTGGATCCGTGCTTCGTCTGCGGGAGGGGCGCTAGCTGCTGAGCCG 480

TgCkUg9-II-5R  AACACCGGCCCATTTGACACCACGCACCAAAAACCGAACCGGTGCCCGATGATGCGCGCC 540
TgCkUg9         AACACCGGCCCATTTGACACCACGCACCAAAAACCGAACCGGTGCCCGATGATGCGCGCC 540

TgCkUg9-II-5R  CAGCTCCAGCAGCGTAGTACTAGAGAACCGCCATGCTTCGTGATCAAACAAATGTCATTC 600
TgCkUg9         CAGCTCCAGCAGCGTAGTACTAGAGAACCGCCATGCTTCGTGATCAAACAAATGTCATTC 600

TgCkUg9-II-5R  GCCGAGAATTTCTATTTGGTACCTGCCCGCGCATTTGAAAGAATTATTACGACATGCGCG 660
TgCkUg9         GCCGAGAATTTCTATTTGGTACCTGCCCGCGCATTTGAAAGAATTATTACGACATGCGCG 660

TgCkUg9-II-5R  TTCTCGGTTGGTGCGGGCAGCTCCAGAATCGCGCTTCACCGAAGGTTCTTCCGAGAGGCA 720
TgCkUg9         TTCTCGGTTGGTGCGGGCAGCTCCAGAATCGCGCTTCACCGAAGGTTCTTCCGAGAGGCA 720

TgCkUg9-II-5R  CAGGATACGGTTTTG 735
TgCkUg9         CAGGATACGGTTTTG 735



TgCkUg9-II-5F  CATCTCAGTGGAACTCGGGGATCACTCTCTGGGACATACTGAAATCAGTTGTGACGAGGC 60
TgCkUg9         CATCTCAGTGGAACTCGGGGATCACTCTCTGGGACATACTGAAATCAGTTGTGACGAGGC 60

TgCkUg9-II-5F  TCCGCTGTTGCCTATAGGGAAACAATGCCGGGATTGTCGCCGTGTTCGCGCTTCAGTGAA 120
TgCkUg9         TCCGCTGTTGCCTATAGGGAAACAATGCCGGGATTGTCGCCGTGTTCGCGCTTCAGTGAA 120

TgCkUg9-II-5F  GCACTTAACCGTGCCGGAGAAATGTGTCGGTGAAGCAATCTCAGACGAGATCCAGGTTTC 180
TgCkUg9         GCACTTAACCGTGCCGGAGAAATGTGTCGGTGAAGCAATCTCAGACGAGATCCAGGTTTC 180

TgCkUg9-II-5F  AAGCGCCCGCAGCTACCAGTGAACATCGCACGTCTGGAAAACAGCGGAGCCGTCTGCTGG 240
TgCkUg9         AAGCGCCCGCAGCTACCAGTGAACATCGCACGTCTGGAAAACAGCGGAGCCGTCTGCTGG 240

TgCkUg9-II-5F  GGTTCTTTTAGAAACGCTGGTACGTGGCGTTTTCAAACAATAACCACACCGCCGGCCCGT 300
TgCkUg9         GGTTCTTTTAGAAACGCTGGTACGTGGCGTTTTCAAACAATAACCACACCGCCGGCCCGT 300

TgCkUg9-II-5F  CTGTTAAAGAACAAAGGTGTGGCACCCAGTTGACCGAAGCGGGAAAGAGTTTCGACGGTG 360
TgCkUg9         CTGTTAAAGAACAAAGGTGTGGCACCCAGTTGACCGAAGCGGGAAAGAGTTTCGACGGTG 360

TgCkUg9-II-5F  TCCCGGTGTCCGACGCTGCCGATGTGGGCATTGCCTGTAACCACGAAGGCGAGAGCGTCC 420
TgCkUg9         TCCCGGTGTCCGACGCTGCCGATGTGGGCATTGCCTGTAACNNNNNNNNNNNNNNNNNNN 420

TgCkUg9-II-5F  ACAGCTGTCTCTGGCGAAGCACCTGATTCTCTAAATAGTTTCTCTGTCACTCAGGCGAGG 480
TgCkUg9         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480

TgCkUg9-II-5F  GCAACGCCGACTTGCCTGTGAGGAACGTGAAGCTATTGTAACTGGAACACCCAGTGAAGT 540
TgCkUg9         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 540

TgCkUg9-II-5F  TAGCCTCCAACAATGTGGCCAGGCATTCGCTGCGGTTTCTTCCGATGGCGGTGTGTGAAC 600
TgCkUg9         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600

TgCkUg9-II-5F  ATGGTAACCGTTCACCGTCGGCGAACTTGCTGTGACTAAAGGAGGCAGGTGTGTCAGCAA 660
TgCkUg9         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660

TgCkUg9-II-5F  ATAGTTCGGTCCCTCGTTGCGTGTCCGCAGGGGAGCCAGGGCTTTTGCCCTCGCTCAACA 720
TgCkUg9         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 720

TgCkUg9-II-5F  ACATCACATCCGTG 734
TgCkUg9         NNNNNNNNNNNNN 734



TgCkUg9-IV-641.m0156 GACTTTGGTTATCAACCCTGGTATCTTGTGAATCGCTTTGCGGGGCTTTGTTGGGGACAC 60
TgCkUg9              GACTTTGGTTATCAACCCTGGTATCTTGTGAATCGCTTTGCGGGGCTTTGTTGGGGACAC 60

TgCkUg9-IV-641.m0156 AACCCAAACGGAACTGCTGCTCTGATTCTGGGAAGTCCGCCGCTGGGATTGTCAGTATAG 120
TgCkUg9              AACCCAAACGGAACTGCTGCTCTGATTCTGGGAAGTCCGCCGCTGGGATTGTCAGTATAG 120

TgCkUg9-IV-641.m0156 AGGCATTGCCAGGCGTCGGCGTTTTCCACCAGCTAATCACAAAATTCGGCAGAATATCGA 180
TgCkUg9              AGGCATTGCCAGGCGTCGGCGTTTTCCACCAGCTAATCACAAAATTCGGCAGAATATCGA 180

TgCkUg9-IV-641.m0156 CAAACTTCTTCTTGCAATCATCCAAGTCTTTGTTCTGTGGGTTGCAGTATTGGGCGCTAT 240
TgCkUg9              CAAACTTCTTCTTGCAATCATCCAAGTCTTTGTTCTGTGGGTTGCAGTATTGGGCGCTAT 240

TgCkUg9-IV-641.m0156 CTGCGGCAGGATTCAAAGTGCCCTCCGACCCACAGTCAATGGTCACCGTGTTCCTTGTTG 300
TgCkUg9              CTGCGGCAGGATTCAAAGTGCCCTCCGACCCACAGTCAATGGTCACCGTGTTCCTTGTTG 300

TgCkUg9-IV-641.m0156 TTGTCATATCCACCTTCAAAGCCTCTGAGTTGCTCTCGGCGCCGTAGGAGCATGTGACGA 360
TgCkUg9              TTGTCATATCCACCTTCAAAGCCTCTGAGTTGCTCTCGGCGCCGTAGGAGCATGTGACGA 360
```

```
TgCkUg9-IV-641.m0156 CGTTCTCGTCTGCAACAGAAGAGGCTCTTGCTTTCACATTGACTGTGAGCTGGCACACTG 420
TgCkUg9              CGTTCTCGTCTGCAACAGAAGAGGCTCTTGCTTTCACATTGACTGTGAGCTGGCACACTG 420

TgCkUg9-IV-641.m0156 CGGGATTTGCAGCTGCTTTGCCGGTGGTGGTGTATGTACACCCGACGTCAAAGGTCTTGT 480
TgCkUg9              CGGGATTTGCAGCTGCTTTGCCGGTGGTGGTGTATGTACACCCGACGTCAAAGGTCTTGT 480

TgCkUg9-IV-641.m0156 CGGATATAGGAAGGTCTTCTTCATTCAGCTTCAGCGTCCACTTTTGAGATTGAATGGCTT 540
TgCkUg9              CGGANNNNNNNNNNNNNNNNNNNNNNNNTCAGCGTCCACTTTTGAGATTGAATGGCTT 540

TgCkUg9-IV-641.m0156 TCTGTTCCGCAGTGGTATTCATCCACTGAACAGCGCTTGTCGTCCCAAGTACGTCTTTCA 600
TgCkUg9              TCTGTTCCGCAGTGGTATTCATCCACTGAACAGCGCTTGTCGTCCCAAGTACGTCTTTCA 600

TgCkUg9-IV-641.m0156 ATGGAACCTGGACGCTCGCCGCGGCGCCTCCGTTTTCCCCAAAATTGCATGGTGTTTCAC 660
TgCkUg9              ATGGAACCTGGACGCTCGCCGCGGCGCCTCCGTTTTCCCCAAAATTGCATGGTGTTTCAC 660

TgCkUg9-IV-641.m0156 TCTGTTTCTGTTTTGTAACACACACATTTCCAGCCAGAGGATTAGGAATCGATGAGATGG 720
TgCkUg9              TCTGTTTCTGTTTTGTAACACACACATTTCCAGCCAGAGGATTAGGAATCGATGAGATGG 720

TgCkUg9-IV-641.m0156 TTCCACCGTCTTTAGCGGTACACTCCAGAGTGGCCGTGAGACTTCCTTTTGACAACGTCA 780
TgCkUg9              TTCCACCGTCTTTAGCGGTACACTCCAGAGTGGCCGTGAGACTTCCTTTTGACAACGTCA 780

TgCkUg9-IV-641.m0156 GGCTCTGTTTACTAACTGCAGTGCTAGATCCCGCGCCTGCCGTAGAAAATTCACACTTCG 840
TgCkUg9              GGCTCTGTTTACTAANNNNNNNNNNNNTCCCGCGCCTGCCGTAGAAAATTCACACTTCG 840

TgCkUg9-IV-641.m0156 CATGTGTCCCCGTGACATCTACCTTGGAACTCACGCCGACTTGAGACTTCAAACTGAGAC 900
TgCkUg9              CATGTGTCCCCGTGACATCTACCTTGGAACTCACGCCGACTTGAGACTTCAAACTGAGAC 900

TgCkUg9-IV-641.m0156 TCTGAAGTCCTTCATGCAGACTGTCGGCAAGACCCTGTCCGCTAGAGAGCAACAAAATTC 960
TgCkUg9              TCTGAAGTCCTTCATGCAGACTGTCGGCAAGACCCTGTCCGCTAGAGAGCAACAAAATTC 960

TgCkUg9-IV-641.m0156 CACCGATACAAACTGCCATCAACTTTCGGACTTGGGACTTGAACCCTCCAAGCCGTTGCT 1020
TgCkUg9              CACCGATACAAACTGCCATCAACTTTCGGACTTGGGACTTGAACCCTCCAAGCCGTTGCT 1020

TgCkUg9-IV-641.m0156 GCATATTGCCTGTCCTT 1037
TgCkUg9              GCATATTGCCTGTCCTT 1037
```

```
TgCkUg9-IX-80.m05038 TTTGTCGAGTATGAAGCGTCGGGTTCGCTTCTTCAACTGCTCCCGGGCGGCACTCGGGTT 60
TgCkUg9              TTTGTCGAGTATGAAGCGTCGGGTTCGCTTCTTCAACTGCTCCCGGGCGGCACTCGGGTT 60

TgCkUg9-IX-80.m05038 GGGGGTCGCGGCTTTGTTCGTCGCTTGTCTCAGTGACATTTCGAGGGCTGTCGCAACAGA 120
TgCkUg9              GGGGGTCGCGGCTTTGTTCGTCGCTTGTCTCAGTGACATTTCGAGGGCTGTCGCAACAGA 120

TgCkUg9-IX-80.m05038 AACGCCGACAATGAGACCCACATGCAAGGAAGAAAAGCAGATGACCACGTGCATGTGCGC 180
TgCkUg9              AACGCCGACAATGAGACCCACATGCAAGGAAGAAAAGCAGATGACCACGTGCATGTGCGC 180

TgCkUg9-IX-80.m05038 AGGTCCTTCTGGTGCAGCCGCGGCAGCGGCCGCGAAAACAATTGCACTTTGTCCCAGTCG 240
TgCkUg9              AGGTCCTTCTGGTGCAGCCGCGGCAGCGGCCGCGAAAACAATTGCACTTTGTCCCAGTCG 240

TgCkUg9-IX-80.m05038 AGTAATGGAATCACTGTTGACTGTTCCACAGATAGTTTCACTGTTGTGCCTTCCGACACA 300
TgCkUg9              AGTAATGGAATCACTGTTGACTGTTCCACAGATAGTTTCACTGTTGTGCCTTCCGACACA 300

TgCkUg9-IX-80.m05038 AGTAAGATTTGCATTGGAGAAGGTCATGACGAAACACTGACGGCGTGTAACAGAAACAAA 360
TgCkUg9              AGTAAGATTTGCATTGGAGAAGGTCATGACGAAACACTGACGGCGTGTAACAGAAACAAA 360

TgCkUg9-IX-80.m05038 GTATCCACAATCAAGGAATTTCTGGGCGTGGCGCACGGCACCAACGCACCACAGTGGACG 420
TgCkUg9              GTATCCACAATCAAGGAATTTCTGGGCGTGGCGCACGGCACCAACGCACCACAGTGGACG 420

TgCkUg9-IX-80.m05038 TCCGACCACGCTCTAACAATTCCAACAGAGCAATTCTCCATAATTGACAAAACCTTATTC 480
TgCkUg9              TCCGACCACGCTCTAACAATTCCAACAGAGCAATTCTCCATAATTGACAAAACCTTATTC 480

TgCkUg9-IX-80.m05038 GCAGGATGCTTAAGAAAAAGTGGGACCAACCAGGGCACGAAGGAGTGCTTGGTGCCTGCC 540
TgCkUg9              GCAGGATGCTTAAGAAAAAGTGGGACCAACCAGGGCACGAAGGAGTGCTTGGTGCCTGCC 540

TgCkUg9-IX-80.m05038 AGTGTCAAAGCCAGGGCCTCGGCAGTGAAGAATGGTGTCCTCATTTGCGCCTACGGCACA 600
TgCkUg9              AGTGTCAAAGCCAGGGCCTCGGCAGTGAAGAATGGTGTCCTCATTTGCGCCTACGGCACA 600

TgCkUg9-IX-80.m05038 GAGAGCAACAATACTGTTTCAGCGGTAACACTGGACGCCAAGAACAACTCTCTGACAATC 660
TgCkUg9              GAGAGCAACAATACTGTTTCAGCGGTAACACTGGACGCCAAGAACAACTCTCTGACAATC 660

TgCkUg9-IX-80.m05038 ATCTGTGACACTGAAGGACAGATGCACCCAGAAAAGGATTCCTTGACTGCTTGCCACTGC 720
TgCkUg9              ATCTGTGACACTGAAGGACAGATGCACCCAGAAAAGGATTCCTTGACTGCTTGCCACTGC 720

TgCkUg9-IX-80.m05038 ACCAATTCCAACACAGAAAACTGTAAGACCTTGAACCTGAGTTCATGCCGAGCTTCACCG 780
TgCkUg9              ACCAATTCCAACACAGAAAACTGTAAGACCTTGAACCTGAGTTCATGCCGAGCTTCACCG 780

TgCkUg9-IX-80.m05038 AGAGGTGGTGGACGCCGGTTGACACAGACAGTAATGCGTTCAAACTCGTGATTCCAGAGG 840
```

```
TgCkUg9                AGAGGTGGTGGACGCCGGTTGACACAGACAGTAATGCGTTCAAACTCGTGATTCCAGAGG 840

TgCkUg9-IX-80.m05038 GAGGCTTTCCTGTTGAGGAAGAAAGGATTGTGTTGGGGTGCACTGAAAACAACAGGGTTC 900
TgCkUg9                GAGGCTTTCCTGTTGAGGAAGAAAGGATTGTGTTGGGGTGCACTGAAAACAACAGGGTTC 900

TgCkUg9-IX-80.m05038 AAGCCTTGAACGGTGACGGTGACAACGACACAGCTGTAAAGCTACCGAC 949
TgCkUg9                AAGCCTTGAACGGTGACGGTGACAACGACACAGCTGTAAAGCTACCGAC 949
```

```
TgCkUg9-XII-65.m0116 ATACTTTCTGCATCATCCCTGCCATCATCCCCACCATCGGCCTCCCCAGAATTCGTGTGG 60
TgCkUg9                ATACTTTCTGCATCATCCCTGCCATCATCCCCACCATCGGCCTCCCCAGAATTCGTGTGG 60

TgCkUg9-XII-65.m0116 TCTGTTGATTCGGATGAGCTTGGAATCGAAGGAGTCGATACGGTGGTTTGTGAATCCGAT 120
TgCkUg9                TCTGTTGATTCGGATGAGCTTGGAATCGAAGGAGTCGATACGGTGGTTTGTGAATCCGAT 120

TgCkUg9-XII-65.m0116 CCGGCGGCAGACACCGTTTTGAACCGCATACAAAAGCTCAACCCTACACTCAAGTACTCT 180
TgCkUg9                CCGGCGGCAGACACCGTTTTGAACCGCATACAAAAGCTCAACCCTACACTCAAGTACTCT 180

TgCkUg9-XII-65.m0116 GCTGGAGGATTTTTAAACGTGTACATCACAACCTTTCCCGCGGAGTCACGTCCTCTTTCA 240
TgCkUg9                GCTGGAGGATTTTTAAACGTGTACATCACAACCTTTCCCGCGGAGTCACGTCCTCTTTCA 240

TgCkUg9-XII-65.m0116 CCAGATTCTGACTCTTCCCGTGTCCACAATTTTGCTGTGTAGCCGGGAAACACATCCTTG 300
TgCkUg9                CCAGATTCTGACTCTTCCCGTGTCCACAATTTTGCTGTGTAGCCGGGAAACACATCCTTG 300

TgCkUg9-XII-65.m0116 AATTGGATTGTCTTTGTGAAATCGCACCCTCCATTCATATATTGATATGCAGCGCTGTAA 360
TgCkUg9                AATTGGATTGTCTTTGTGAAATCGCACCCTCCATTCATATATTGATATGCAGCGCTGTAA 360

TgCkUg9-XII-65.m0116 GCGTCAGTCAATGTTCCAGATGAATTCTCGCTAGTTTCGGGTTGATAAACGGCACTCGCT 420
TgCkUg9                GCGTCAGTCAATGTTCCAGATGAATTCTCGCTAGTTTCGGGTTGATAAACGGCACTCGCT 420

TgCkUg9-XII-65.m0116 GTTTCGTCAGTGACCCGGATCGTGCCAGAAGATCCAAGAGAAAAGACTTCTTCAACATCT 480
TgCkUg9                GTTTCGTCAGTGNNNNNGATCGTGCCAGAAGATCCAAGAGAAAAGACTTCTTCAACATCT 480

TgCkUg9-XII-65.m0116 CCCTCCAATCCCCCTTTCGGAATGATGGTGCTAAAATCAACCTCAGCTGTCGGTTGACCG 540
TgCkUg9                CCCTCCAATCCCCCTTTCGGAATGATGGTGCTAAAATCAACCTCAGCTGTCGGTTGACCG 540

TgCkUg9-XII-65.m0116 GATTCGGCGGCGACGCACTGCAAGACAGTGCCAATGAAAAGGACTGAACACCCGCCGCGA 600
TgCkUg9                GATTCGGCGGCGACGCACTGCAAGACAGTGCCAATGAAAAGGACTGAACACCCGCCGCGA 600

TgCkUg9-XII-65.m0116 AAGATCTGCCTTGCCATTGCGAGTGCGGATC 631
TgCkUg9                AAGATCTGCCTTGCCATTGCGAGTGCGGATC 631
```

```
TgCkUg9-X-33.m02185  TGCTGCTCATTACCGTCGGATCGCTGCTGACCGCTTCCGAATCGGTCCAGCTTTCTGAAG 60
TgCkUg9                TGCTGCTCATTACCGTCGGATCGCTGCTGACCGCTTCCGAATCGGTCCAGCTTTCTGAAG 60

TgCkUg9-X-33.m02185  GCATGAAGAGGCTCAGCATGAGAGGAAGAAGTCCCAGCCCGAAAACGGGAAGATTTGAGA 120
TgCkUg9                GCATGAAGAGGCTCAGCATGAGAGGAAGAAGTCCCAGCCCGAAAACGGGAAGATTTGAGA 120

TgCkUg9-X-33.m02185  GTGGAGACGAAGGGACATCTACAATGTCGCCCTCCGTCGCAGCACGCCAACAGGAACTAG 180
TgCkUg9                GTGGAGACGAAGGGACATCTACAATGTCGCCCTCCGTCGCAGCACGCCAACAGGAACTAG 180

TgCkUg9-X-33.m02185  GGCTGCTCCGTCCGGAAGAGAGGCTCATTGCGGGTCAAGCCAAAGCAGCGGCGCTACAGA 240
TgCkUg9                GGCTGCTCCGTCCGGAAGAGAGGCTCATTGCGGGTCAAGCCAAAGCAGCGGCGCTACAGA 240

TgCkUg9-X-33.m02185  CTGTCCATCAATTAGGAGCAGTTGTTCTTACCCCAGAGCAAGCTAAAGCCGCATTGCTAG 300
TgCkUg9                CTGTCCATCAATTAGGAGCAGTTGTTCTTACCCCAGAGCAAGCTAAAGCCGCATTGCTAG 300

TgCkUg9-X-33.m02185  ACGAAATTCTCAGAGCCACGCAAAATTTGGACCTCAAGAAGTACGAGAATCTCAACACGG 360
TgCkUg9                ACGAAATTCTCAGAGCCACGCAAAATTTGGACCTCAAGAAGTACGAGAATCTCAACACGG 360

TgCkUg9-X-33.m02185  AGCAACAAAAGGCCTATGAACAAGTCCAGAAGGATCTTTCGCTACTGAGTCCGGAAACGA 420
TgCkUg9                AGCAACAAAAGGCCTATGAACAAGTCCAGAAGGATCTTTCGCTACTGAGTCCGGAAACGA 420

TgCkUg9-X-33.m02185  AGGCTTTGCTCATTGAAAATCATAGGAAGGAGAAATCCCTACTTGATCAGGCGAAAAGGC 480
TgCkUg9                AGGCTTTGCTCATTGAAAATCATAGGAAGGAGAAATCCCTACTTGATCAGGCGAAAAGGC 480

TgCkUg9-X-33.m02185  TGTTCCGAAGACGCCACTACCACGTTACCAGGCAGGCGGCTTTGGCTGGACAAATCCTCA 540
TgCkUg9                TGTTCCGAAGACGCCACTACCACGTTACCAGGCAGGCGGCTTTGGCTGGACAAATCCTCA 540

TgCkUg9-X-33.m02185  ACGAGCAGAGGGATGCGAGTGGAGCATTACAATCAGGTGCTGTGAAGGCTGCGATCCGAA 600
TgCkUg9                ACGAGCAGAGGGATGCGAGTGGAGCATTACAATCAGGTGCTGTGAAGGCTGCGATCCGAA 600

TgCkUg9-X-33.m02185  AAGCCAACGAGCAGTACAACGTCGCTGAAGAAGACAAGAACTTTAACGAGGAACAGCATG 660
TgCkUg9                AAGCCAACGAGCAGTACAACGTCGCTGAAGAAGACAAGAACTTTAACGAGGAACAGCATG 660
```

```
TgCkUg9-X-33.m02185  CAGCACAACTGAAAAAGGTAGGCGCTATGCCCTAGTCGTAAAAACGACGACGAACCCGCG 720
TgCkUg9              CAGCACAACTGAAAAAGGTAGGCGCTATGCCCTAGTCGTAAAAACGACGACGAACCCGCG 720

TgCkUg9-X-33.m02185  TCGGAAATCCCGAATCTGTTTGGGATGGCTTTGACCGGTATGAAGAGGTCAAACACAGCT 780
TgCkUg9              TCGGAAATCCCGAATCTGTTTGGGATGGCTTTGACCGGTATGAAGAGGTCAAACACAGCT 780

TgCkUg9-X-33.m02185  TCTTGTCAGACCAACATTGGTCAATTCTGATTTTTTCCGTTTATTCGTTTTCCATGATGA 840
TgCkUg9              TCTTGTCAGACCAACATTGGTCAATTCTGATTTTTTCCGTTTATTCGTTTTCCATGATGA 840

TgCkUg9-X-33.m02185  ACAATCACTTTGAATATGGTCTGCCGCTCAGCACGAA 877
TgCkUg9              ACAATCACTTTGAATATGGTCTGCCGCTCAGCACGAA 877



TgCkUg9-VIIa-20.m038 CGGTCTCCAAGGCTTGTAGCGGGAGCATCCGTTTCTGAGGCTCTCGATTCAAGAGGTTTC 60
TgCkUg9              CGGTCTCCAAGGCTTGTAGCGGGAGCATCCGTTTCTGAGGCTCTCGATTCAAGAGGTTTC 60

TgCkUg9-VIIa-20.m038 GGATGAGGTCTTGAACCAGCTCAGGCGTGGAAGGACAATCTGCGAAGTGTAAGTAGTCCC 120
TgCkUg9              GGATGAGGTCTTGAACCAGCTCAGGCGTGGAAGGACAATCTGCGAAGTGTAAGTAGTCCC 120

TgCkUg9-VIIa-20.m038 AGATGCCGTCGGCCGGAGTTGGACGTTCCTTGCACCAGATGCAGTACAAAGTTATACCGA 180
TgCkUg9              AGATGCCGTCGGCCGGAGTTGGACGTTCCTTGCACCAGATGCAGTACAAAGTTATACCGA 180

TgCkUg9-VIIa-20.m038 GTTGCCACGCGTCAGTGGTGAATGTATAGGTGATGTCTGTAGTCTGAAACGGTCGCTCCG 240
TgCkUg9              GTTGCCACGCGTCAGTGGTGAATGTATAGGTGATGTCTGTAGTCTGAAACGGTCGCTCCG 240

TgCkUg9-VIIa-20.m038 GAGGCTCGTAACCGGGAGTACCTATCGCGGGTCCAACCGAATTATTGATTCTATACGTTC 300
TgCkUg9              GAGGCTCGTAACCGGGAGTACCTATCGCGGGTCCAACCGAATTATTGATTCTATACGTTC 300

TgCkUg9-VIIa-20.m038 CGAAGTCGCCGAGAAACAGGCGACCGTCTTTCAAGAGGAGGAAATTCGCCGGTTTGATAT 360
TgCkUg9              CGAAGTCGCCGAGAAACAGGCGACCGTCTTTCAAGAGGAGGAAATTCGCCGGTTTGATAT 360

TgCkUg9-VIIa-20.m038 CCGTATGCACAATTCCCTGAGCTTGAACATTGGCCACTAGTCTGATTGCCTGACTGGATA 420
TgCkUg9              CCGTATGCACAATTCCCTGAGCTTGAACATTGGCCACTAGTCTGATTGCCTGACTGGATA 420

TgCkUg9-VIIa-20.m038 GGTACATTCGAACGACCAGGCCTAATTCACTGTTATTGACAGACGCATCTCCAAATACCC 480
TgCkUg9              GGTACATTCGAACGACCAGGCCTAATTCACTGTTATTGACAGACGCATCTCCAAATACCC 480

TgCkUg9-VIIa-20.m038 ATGAAATGACTTTGCTCATGTCCGTTTCTGCCCGCATCATGAGAAGAAAATAGTTTGGTA 540
TgCkUg9              ATGANNNNNNNNNNNNNNNNCCGTTTCTGCCCGCATCATGAGAAGAAAATAGTTTGGTA 540

TgCkUg9-VIIa-20.m038 CCCACCGAGTCGTCAAACCAATCACGACCTCTGTGGATGCTGGCTGTCCCTCTAACATCA 600
TgCkUg9              CCCACCGAGTCGTCAAACCAATCACGACCTCTGTGGATGCTGGCTGTCCCTCTAACATCA 600

TgCkUg9-VIIa-20.m038 CAACATCACTAGGAACCATGAATCTACAGCGTTCCTGGGCATCCTTCGCCGTCTTGGCTA 660
TgCkUg9              CAACATCACTAGGAACCATGAATCTACAGCGTTCCTGGGCATCCTTCGCCGTCTTGGCTA 660

TgCkUg9-VIIa-20.m038 GACTAAAGTTCCTGTAGCAGAACGACTCCCTCTGCAAGTCACGCATAGTCTCATCGGTGG 720
TgCkUg9              GACTAAAGTTCCTGTAGCAGAACGACTCCCTCTGCAAGTCACGCATAGTCTCATCGGTGG 720

TgCkUg9-VIIa-20.m038 GCTCCTTTTCTGACATGAAAACCTTAACAGCCAACTCTTCATTCGTCTCCACGTCTGTAG 780
TgCkUg9              GCTCCTTTTCTGACATGAAAACCTTAACAGCCAACTCTTCATTCGTCTCCACGTCTGTAG 780

TgCkUg9-VIIa-20.m038 CCTCATATACAGTGGCGAATCCACCAGAGCCAAGGGGTGCACCTCTCACCAGCGTCCGGG 840
TgCkUg9              CCTCATATACAGTGGCGAATCCACCAGAGCCAAGGGGTGCACCTCTCACCAGCGTCCGGG 840

TgCkUg9-VIIa-20.m038 TTTCACCGGTAGTCACAGAAACAAATCTTTCTGGAACGTCCGGAGGCCAGGCAGCGACAG 900
TgCkUg9              TTTCACCGGTAGTCACAGAAACAAATCTTTCTGGAACGTCCGGAGGCCAGGCAGCGACAG 900

TgCkUg9-VIIa-20.m038 TCAGAATTCTATCCAGTCGTTCACTATTTCCTAGCGCTTGAGGCTGCCCGATTTGTTCTT 960
TgCkUg9              TCAGAATTCTATCCAGTCGTTCACTATTTCCTAGCGCTTGAGGCTGCCCGATTTGTTCTT 960

TgCkUg9-VIIa-20.m038 GCATGTGCGCCAGGATGCGTTTGCCGATGACGCATCCAGAATCCGCCTTCTCAAAAAACA 1020
TgCkUg9              GCATGTGCGCCAGGATGCGTTTGCCGATGACGCATCCAGAATCCGCCTTCTCAAAAAACA 1020

TgCkUg9-VIIa-20.m038 ATTCAGATCGCCGTCTTTGTGCCCGCAAACTTCGCAGTCGGTTACGCCCTTGAGAAAAGT 1080
TgCkUg9              ATTCAGATCGCCGTCTTTGTGCCCGCAAACTTCGCAGTCGGTTACGCCCTTGAGAAAAGT 1080

TgCkUg9-VIIa-20.m038 ACCTCGGGATTCCGCGTGTGAAGAACGATTTCGCTTTACGCCAAAGTTGAGCAAGCCTCT 1140
TgCkUg9              ACCTCGGGATTCCGCGTGTGAAGAACGATTTCGCTTTACGCCAAAGTTGAGCAAGCCTCT 1140

TgCkUg9-VIIa-20.m038 GTCGTACGGACAGACGGGGTCTCGAGGAGCTATCAGCTGCGGAACCCGAGACTTCACCAT 1200
TgCkUg9              GTCGTACGGACAGACGGGGTCTCGAGGAGCTATCAGCTGCGGAACCCGAGACTTCACCAT 1200

TgCkUg9-VIIa-20.m038 CTCTTTCTCTCCGTCTCAAGAGATTCAGAAGCCTCTGAAAAATCGTGCGCCGTTGGTTTG 1260
TgCkUg9              CTCTTTCTCTCCGTCTCAAGAGATTCAGAAGCCTCTGAAAAATCGTGCGCCGTTGGTTTG 1260

TgCkUg9-VIIa-20.m038 TTTCTTGCGCATTTACCCGGTGTTGAGCCCTTTCCTCCAGCAATGAAACGTCTCGAGTGC 1320
```

```
TgCkUg9               TTTCTTGCGCATTTACCCGGTGTTGAGCCCTTTCCTCCAGCAATGAAACGTCTCGAGTGC 1320

TgCkUg9-VIIa-20.m038 TTTCTGTCGCTCCTGCTAAATATTTGTAGTGGCCAACTGTAGCAGGCCACCGCTTGTCAG 1380
TgCkUg9               TTTCTGTCGCTCCTGCTAAATATTTGTAGTGGCCAACTGTAGCAGGCCACCGCTTGTCAG 1380

TgCkUg9-VIIa-20.m038 CAACGTGCTGTTGCGAATCCAAACTTGTCGGTTTGGAGTCGAGTTTTGAAGGATCAAGTG 1440
TgCkUg9               CAACGTGCTGTTGCGAATCCAAACTTGTCGGTTTGGAGTCGAGTTTTGAAGGATCAAGTG 1440

TgCkUg9-VIIa-20.m038 TGATTCCGGTCCCATCCTGGACTTGGAAGAGCAGGAAGACAAGCGCTACATTTAACACCG 1500
TgCkUg9               TGATTCCGGTCCCATCCTGGACTTGGAAGAGCAGGAAGACAAGCGCTACATTTAACACCG 1500

TgCkUg9-VIIa-20.m038 CAAGACAGGCTGTCTTCGGGAGAAGAGTCGCTAAACCCATTCGGACGACGGTACGCGTAA 1560
TgCkUg9               CAAGACAGGCTGTCTTCGGGAGAAGAGTCGCTAAACCCATTCGGACGACGGTACGCGTAA 1560

TgCkUg9-VIIa-20.m038 GAGGTGGCCGCTGTACCGAAAACATCACAACTTTCACACAA 1601
TgCkUg9               GAGGTGGCCGCTGTACCGAAAACATCACAACTTTCACACAA 1601
```